nature

AVIAN
EVOLUTION
A dinosaur in
the hand

NEW UNITS
FOR OLD
Atomic for
our times

QUANTUM
MECHANICS
A superconducting
qubit measured

SWINE FLU
SO FAR

The emergence of
pandemic H1N1

Abstractions

SECOND AUTHOR
Modern-day birds are widely thought to have descended from dinosaurs, but a nagging problem has been muddying the waters. The three-fingered 'hands' of the theropod dinosaurs — considered to be birds' closest relatives — were thought to have developed by retaining the first, second and third digits of a five-fingered ancestral hand belonging to an earlier species of dinosaur. Meanwhile, the 'hands' of modern birds — now much smaller than when they first evolved, and embedded in the wing — are thought to have retained the second, third and fourth digits of the ancestral hand. Palaeontologist James Clark of George Washington University in Washington DC tells *Nature* how he, lead author Xing Xu and their team linked the two hands (see page 940).

How did you connect birds' hands to those of theropods?
We found a transitional species of theropod at the point in this group's evolution that it underwent a major reduction in its number of fingers. If you look at the hand of this species, *Limusaurus inextricabilis*, you can see it's in the process of losing its innermost, or first, finger. The nubbin of this digit is visible next to the second and third fingers. This contradicts the long-standing theory that the first finger was retained during evolution, and that the outer fingers disappeared. It was the puzzle piece we needed.

Where did you find this species?
In the Shishugou Formation in Xinjiang, a region in the northwestern part of China. In 2001, we started discovering dinosaur 'death pits' — former mud pits that had solidified — in the Wucaiwan region. It's beautiful countryside — parts of the movie *Crouching Tiger, Hidden Dragon* were filmed there. You wouldn't know it had once been mud pits.

What did you uncover in the death pits?
We found stacks of dinosaur skeletons interspersed with layers of rocks. The third pit we came upon contained 16 animals, including several intact theropod skeletons. One was the most complete we'd ever found: it was the first to contain the skull, the shoulder and the arms. And it had an articulated hand. The hand bones, which were mostly preserved, provided the first clue to its transitional status.

Is this sort of work physically demanding?
It can be. We were camping out in rugged desert terrain with sandstorms and torrential rains that formed a river through our camp. It's usually around 37 °C in Wucaiwan, but one area we worked in was pushing the high 40s. We had to get our shovels out to dig a road so that we could bring the fossil blocks back in the truck to be shipped.

MAKING THE PAPER
Ian Harding & James Eldrett

Pollen fossils provide a timeline for cooling in Greenland.

Ian Harding (left) and James Eldrett.

Some 33.5 million years ago, a warm, humid greenhouse Earth gave way to a cooler planet that would eventually have ice blanketing its poles. The change, one of Earth's most profound climate shifts, marks the boundary between the Eocene and Oligocene epochs. However, during the past few years there has been some debate about when the first polar ice actually formed.

Two years ago, Ian Harding, James Eldrett and their colleagues reported physical evidence that ice was present during the Eocene–Oligocene transition in the region of modern-day Greenland, and that the ice may have formed as early as 38 million years ago — earlier than expected (J. Eldrett *et al. Nature* **446**, 176–179; 2007). The finding caused a stir because atmospheric carbon dioxide concentrations for the period had been estimated at double pre-industrial levels, meaning that it was unlikely that they could have supported the development of large ice sheets.

Harding, Eldrett and two new co-authors have now gathered sufficient evidence to reconstruct the climatic conditions in Greenland during the Eocene–Oligocene transition. They find that temperatures would have allowed the build-up of ice to begin during that time, putting the argument to rest.

The project first started in 2003, when Eldrett was finishing up his doctoral work in Harding's lab at the National Oceanography Centre in Southampton, UK. He was analysing a sediment core from a drilling site deep below the Norwegian–Greenland Sea, which turned out to be one of the most complete geological records of the Eocene–Oligocene boundary. An analysis of pebble striations in one sample led to the 2007 study that revealed the earliest evidence of ice on Greenland. For Eldrett, it was also the beginning of a stint of moonlighting, following up on the finding at nights and over weekends while pursuing a day job as a stratigrapher for Shell Exploration and Production UK in Aberdeen.

Eldrett and Harding had not previously been able to obtain climatic data from samples from the Norwegian-Greenland Sea site, which would support their 2007 finding of ice formation during the Eocene–Oligocene transition. This was because the samples were completely devoid of biogenic calcium carbonate, a chemical signature typically used to reconstruct past climates. The samples were, however, incredibly rich in well-preserved microfossils of marine plankton, and of pollen and spores that had been washed into the sea from terrestrial plants. "Looking down the microscope for the first time, there were pristine examples of the diverse, organic-walled fossils characteristic of that time period," recalls Eldrett.

> "The climate shift coincides exactly with the first appearance of ice on Greenland."

The duo figured that they could capitalize on the pollen fossil record to track changes in vegetation and, by proxy, climate changes. "We knew which plant species were present, but we didn't know how to translate that into climate changes," says Eldrett. "So our first port of call was David Greenwood, to try to get him on board." Greenwood, a palaeobotanist at Brandon University in Manitoba, Canada, had previously reconstructed past climates in the Arctic Circle on the basis of plant fossils.

To link the fossils to climate, the team opted for a method that uses the nearest living relative of an ancient plant species to define the temperature and precipitation parameters likely to support that plant's growth. Their analysis reveals that during the early Eocene, Greenland's flora was similar to that of modern-day Florida, and included ferns, palms and cycads, as well as trees such as hickory and willow. Later Eocene sediments also contained fir, spruce and pine pollens from coniferous forests. In the Eocene–Oligocene boundary sediments, coniferous trees dominate, with frost-sensitive plants such as palms disappearing.

The team's climate reconstruction indicates that mean winter temperatures in Eocene Greenland were above 5 °C, whereas during the Eocene–Oligocene transition they dropped to 0–2 °C. As the Oligocene approached, Greenland began experiencing colder winters and greater annual temperature ranges (see page 969).

The climate shift coincides exactly with the first appearance of ice on Greenland, indicating that, even with higher CO_2 levels, the initiation of continental ice formation in polar regions may at least have been possible. "This work is another page in the story of unravelling the driving mechanisms of profound climate change in the geological past," says Harding.

With this study completed, Eldrett says he'll probably slow his moonlighting down. This summer, he plans to spend his weekends walking in the Scottish Highlands and cycling. ■
See also *Nature* 446, xiii; 2007.

Animal farm: pig in the middle

The 2009 flu pandemic highlights the urgent need for an independent international body for research into human diseases that originate in animals.

When animal pathogens make the leap into humans — as has happened with the 2009 pandemic virus that originated in swine — animal-health scientists can find themselves in an awkward position. Unlike their colleagues in public health, who focus their energies on protecting the planet's 6.8 billion humans, animal-health specialists tend to work through government agencies, whose primary mission is to promote and protect national and international livestock and meat trade.

This focus on commerce can sometimes lead to conflicts of interest, as well as some policy positions that border on denial. Since the first outbreaks of the 2009 pandemic virus in the United States and Mexico, for example, the Paris-based World Organisation for Animal Health (OIE) has expended considerable energy trying to keep people from calling the virus 'swine flu'. The OIE's quite legitimate concern is that this nomenclature might adversely affect trade, with countries taking unnecessary measures such as culling herds, or invoking trade bans on pigs and pork. From a strictly scientific point of view, however, there is abundant genetic evidence that the name is appropriate. It is a reassorted swine influenza virus that has jumped from pigs to humans.

The OIE has also played down the possibility that the 2009 pandemic flu might be spreading in pigs, noting that it has not been found in any animals outside of one farm in Canada. But how vigorous has the search been? There is no requirement that the authorities be notified of flu in pigs, as the animals generally recover, and farmers have little incentive to report an outbreak in their herds given the potential repercussions. Furthermore, little funding has been available for extensive surveillance. A case in point is the European Surveillance Network for Influenza in Pigs, whose paltry €100,000 (US$139,000) in annual funding expired in March, just a month before the pandemic strain was first detected. Yet public-health researchers say that if the virus is circulating in pigs, and moving back and forth between pigs and humans, it increases the risk that the virus will genetically reassort into a more dangerous pathogen (see page 894).

The human–animal disease interface is fraught with such competing agendas. But to the OIE's credit, it has had a key role in creating a body that could be a model for a credible, honest broker. Founded jointly with the United Nations' Food and Agriculture Organization (FAO) in 2005, the OIE/FAO Network of Expertise on Animal Influenza (OFFLU) has been bringing together labs working on surveillance and research of human infectious diseases that have arisen in animals. OFFLU has also been outspoken on the need for countries to share virus samples and sequences for research (see *Nature* **440**, 255–256; 2006) and has built important bridges with the World Health Organization (WHO) and other public-health agencies.

> "Surveillance of human diseases that originate in animals remains in the nineteenth century."

What is needed now is international support for a greatly expanded OFFLU-like network that has enough funding to do its own research and to coordinate global surveillance efforts on influenza and other diseases emerging from animals. The WHO and other public-health organizations should also be made an integral part of the network.

The 2009 pandemic has forced scientists to confront the elephant — or pig — in the room, which is that surveillance of human diseases that originate in animals remains in the nineteenth century (see *Nature* **440**, 6–7; 2006), and is chronically underfunded. Animal- and public-health bodies must now step up and fund a serious joint initiative in this area. ∎

Coherent advocacy please

Reactions to UK government changes are an example of how researchers should not behave in a downturn.

As high-energy physicists and astronomers learned long ago, when making the case for investments in science it is helpful or even essential to present a coherent front to outsiders, not least governments. The idea that the entirety of scientific research could present such a united front in public will strike most of *Nature*'s readers as fanciful. Yet some coherence will be needed in the months and years ahead if science is to maintain the public support it needs, at a time when severe economic and other challenges assail scientifically active countries.

Most politicians appreciate the bounties of science; for example, drugs now entering clinical trials that are direct outcomes of basic biomedical research, and experimental revelations about the Universe and human origins. But they do not necessarily understand the process of science — the unpredictabilities of fundamental research, the uncertainties in applying that research to real problems, and the sheer scale of effort required to make headway.

In straitened times, the politicians within any government or legislature who do understand the importance of science need all the help they can get from the research community as they seek to maintain its financial support. How scientists should not react in such an environment has been well illustrated over the past two weeks in the United Kingdom. The Labour government, weakened by resignations and recent election losses, instituted a reshuffle of ministers and ministries that moved funding for science and universities away from a dedicated department created two years ago and into the new Department for Business, Innovation and Skills. Immediately there

came laments from researchers that science would now be the servant of industry and that fundamental research might lose out.

Such reactions are misguided, both as a matter of fact and in principle. Factually, they ignore the track record of the current government, whatever its prevailing structure, in its consistent support for a portfolio of research ranging from the fundamental to the directly applicable. They also do insufficient justice to the two ministers now at the helm of the university and science bases, both of whom are strong advocates for science. The science minister Paul Drayson (who has been given more clout across government than his predecessors in the role) is an articulate businessman who has never lost sight of the fundamental science on which his pharmaceutical company PowderJect was based. And his ministerial boss, Peter Mandelson, is a highly effective politician who is pivotal within the government, and who previously demonstrated an understanding of the needs of science when, in 1998, he was Secretary of State for Trade and Industry and in charge of the government department that was then relevant to science funding.

The reactions to the reshuffle were also misguided in principle. Whether dealing with the Labour government or the Conservative opposition, UK scientists as a whole need to avoid giving the impression that they are impervious to the requirements of the nation and that any outsider should simply give them the money and leave them to get on with it. This will be especially true over the next 12 months, as the country heads towards a general election and as both main political parties plan future expenditures.

None of this is intended to let politicians off the hook. The government needs to articulate more powerfully its vision for the universities, and the Conservative party needs to explain its plan for how science and the universities will be kept robust in the difficult times ahead.

But the science community in any country where national budgets are under extraordinary pressure should be sending coherent positive messages to all political parties. These communities should find fresh language with which to extol the value of research in meeting national challenges. They need to highlight its relevance to a nation's particular economic opportunities, to mitigating and anticipating such effects as climate change and emerging diseases and, yes, to sustaining its vigour and enriching the nation's culture, through fundamental new insights. If, by contrast, researchers and their supporters convey to the public a sense of entitlement, they risk undermining science as a whole. ■

Degrees of knowledge

Technology is founded on precision measurements, and scientists strive to make these ever more exact.

For the nineteenth-century British physicist William Thomson, later Lord Kelvin, the act of measurement was at the heart of science. "Nearly all the grandest discoveries of science have been but the rewards of accurate measurement," he once declared.

It was thus fitting that, in 1954, Kelvin was honoured for his own pioneering work in thermodynamics when the fundamental unit of temperature was named after him. Today, the kelvin stands ready to be redefined (see page 902), thanks to the efforts of researchers who have taken Kelvin's advice to heart and pushed the bounds of precision measurements in ways that he would not have dreamed possible. Working at national standards laboratories in several countries, these scientists do not usually garner the kinds of accolade bestowed on leaders in more prominent fields. But science as a whole should celebrate the work that today's precision metrologists and their forebears have accomplished.

Centuries ago, metrology played a more obvious role in science and in society in general. The Royal Observatory in London and the French Royal Academy of Sciences in Paris were established in the latter part of the seventeenth century in part to pin down measurements of longitude, which was of tremendous practical importance for international commerce. A century later, efforts to standardize the units of measurement led to the adoption of the metric system in France, which replaced units based on inexact anatomical references, such as the foot. That system was eventually enshrined in 1875 by a diplomatic treaty known as the Metre Convention.

By the early twentieth century, countries had developed laboratories devoted to precision metrology, such as the National Physical Laboratory in Teddington, UK, and the US National Bureau of Standards, now renamed the National Institute of Standards and Technology (NIST) and located in Gaithersburg, Maryland. The renaming was not only symbolic; the pursuit of exacting measurement continues to play an important part in the development and proliferation of new technologies. Today, the forefront of metrology enables advances in such varied fields as the semiconductor industry, climate change, biomedical science and the burgeoning area of nanotechnology. At the same time, national metrological labs also carry out important work in basic science. NIST researchers, for example, have shared physics Nobels for the development of laser-based precision spectroscopy, Doppler cooling and the creation of the first Bose–Einstein condensate.

From the vantage point of the average citizen or scientist, the value of refining the units of measurement is not as obvious today as when nations fought over the definition of units in the trade of grain or cloth. We are now in the 'long tail' of improvements, making the tiniest of adjustments to accuracy and precision, and fiddling with subtleties in definition that largely improve the aesthetics of our system of measurement rather than its immediate utility.

But there is intrinsic value in pushing nature to the absolute limits of experimental precision and seeing whether our current understanding cracks. It was precision measurement of the density of nitrogen, for example, that led to the discovery of the element argon. Lord Kelvin, in his presidential address to Britain's Royal Society in 1894, rightly acknowledged that work as the most important scientific discovery of the year. In his address, he referred to something he had said two decades earlier: "Accurate and minute measurement seems to the non-scientific imagination a less lofty and dignified work than looking for something new." But, as today's precision metrologists have shown, many discoveries will no doubt arise from the ongoing desire to express our knowledge in numbers. ■

RESEARCH HIGHLIGHTS

A grey's anatomy

Cell **137,** 1088–1099 (2009)

It seems that stress really can turn hair grey. Emi Nishimura of Tokyo Medical and Dental University and her colleagues have found that ionizing radiation turns mouse fur permanently grey by depleting the stem cells that make melanocytes — pigment-producing cells — in the hair follicle. DNA-damaging chemicals have a similar effect.

Unexpectedly, cell death was not directly to blame for this stem-cell depletion. Rather, DNA damage prolonged the activation of a protein called MITF, which regulates melanocyte maturation. This activation enhanced the cells' development into melanocytes, leaving no stem cells behind to colour the next cycle of hair growth.

AGRICULTURE
No hectare spare

Glob. Change Biol. **15,** 1716–1726 (2009)

Increased agricultural yields do not necessarily free up land for nature, say Robert Ewers, now at Imperial College London, and his colleagues. Some conservationists have argued that intensive agriculture could result in such 'land sparing', but other forces tend to keep land under the plough.

The researchers analysed changes in yields of the world's 23 most important food crops and in per-capita cropland area for 124 countries between 1979 and 1999. Developing countries showed the strongest link between increased yields and reduced farmland area, but this was largely counteracted by farmers growing cash crops on land no longer needed for staples. No such trend was seen for developed countries — perhaps, the team suggests, because agricultural subsidies mean that land is farmed whether it's needed or not.

HUMAN EVOLUTION
Not so clean sweep

PLoS Genet. **5,** e1000500 (2009)

Neutral forces such as population migrations and the random accumulation of mutations seem to have had an important role in shaping how natural selection acts on human genes, observe Jonathan Pritchard of the University of Chicago in Illinois and his colleagues.

The team compared the relative occurrence of variations at individual nucleotides among the genomes of almost 1,000 individuals from diverse geographic locations to see in what cases variations have become fixed.

As expected, variations in and around genes showed stronger signs of fixation, indicative of selection. But only rarely did strong selective action allow a new mutation to sweep through the population and become fixed. The authors suggest that simultaneous partial sweeps may have driven adaptation, or that humans are less adaptable than many suspected.

PHYSIOLOGY
Pain's new path

Cell **137,** 1148–1159 (2009)

Delta and mu opioid receptors regulate pain transmission in the nervous system. Morphine and some endogenous opioids act on mu receptors to relieve pain, for example. But the role of delta receptors has been less clear.

The two types of receptor have long been thought to occupy the same pain-sensing neurons and to work in a cooperative fashion. But by tracking the positions of delta opioid receptors fused to green fluorescent protein in mice, Allan Basbaum of the University of California, San Francisco, and his colleagues suggest that several assumptions made about this receptor are untrue. Mu and delta opioid receptors inhabit different nerve fibres, and have separate roles in controlling pain: mu receptors for heat and delta receptors for mechanical stimuli.

PLANT GENETICS
Asexual *Arabidopsis*

PLoS Biol. **7,** e1000124 (2009)

Plant breeders have come a step closer to apomixis — asexual production of seeds — in plants that normally reproduce sexually. The switch could lead to consistently higher yields in some crops by stemming the genetic mixing that occurs during sexual reproduction.

Raphaël Mercier of the French National Institute for Agricultural Research in Versailles and his colleagues identified an *Arabidopsis thaliana* gene in which mutation prevents plants from completing the second round of cell division that occurs during meiosis, the sexual form of cell division that produces pollen and ova. They call this gene omission of second division (*OSD1*). When the researchers combined mutations in *OSD1* with mutations in previously discovered meiosis genes, the resulting plants produced their reproductive cells by asexual division, or mitosis.

For a longer version of this story, see: http://tinyurl.com/mk2xe8

IMMUNOLOGY
Trouble for termites

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0904063106 (2009)

Termites and other insect pests could be vanquished by short-circuiting their immunity and letting microbial infections do the rest.

Mark Bulmer of Northeastern University in Boston, Massachusetts, Ram Sasisekharan at the Massachusetts Institute of Technology in Cambridge and their colleagues studied a Gram-negative-bacteria-binding protein, GNBP2, that termites maintain in immune cells and secrete throughout their nests. The protein recognizes carbohydrates associated with infectious microbes, and cleaves them, creating fragments that prime the insects' immune systems.

K. INOMATA
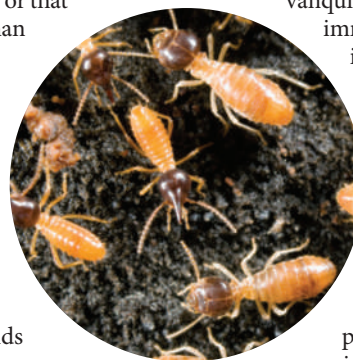
S. STAMMERS/SPL

The researchers identified a glucose derivative that binds to an active pocket of GNBP2 and disarms it. This molecule, D-δ-gluconolactone, caused termites to succumb to subsequent infection with the fungus *Metarhizium anisopliae* as well as opportunistic infections in lab tests.

QUANTUM MECHANICS
Do the wave–particle

Nature Phys. doi:10.1038/nphys1278 (2009)

In the whacky world of quantum mechanics, particles can act as waves and waves as particles. Physicists wondered whether this duality might also be true for a form of collective motion known as a 'surface plasmon polariton', which arises when light excites electrons on a metallic surface.

Among other quantum properties, plasmon polaritons seem to have wave–particle behaviour. Fedor Jelezko and Jörg Wrachtrup at the University of Stuttgart in Germany and their colleagues looked at plasma polaritons on the surface of a silver nanowire and found that they interfered with themselves — a telltale signature of wave–particle duality. The authors say the work will prove useful, particularly in the development of quantum networks.

CHEMISTRY
Don't be square

*Angew. Chem. Int. Edn* **48,** 4824–4827 (2009)

Tiny silver crystals, which among other things are useful as catalysts and in biomedical imaging, like to form as cubes. But to allow better control of their properties, these crystals need to be coaxed into different shapes.

Younan Xia at Washington University in St Louis and his colleagues made silver nanocubes with a conventional reaction using silver nitrate solution, but then quickly added a small, additional amount of this solution. Three of the cubes' faces were augmented with extra silver atoms, creating a crystal that is half octahedron and half truncated cube (pictured below). With slower addition of silver nitrate, the growing cubes remained cubic.

The authors suggest that the exotic crystal shapes could prove useful for an imaging technique called surface-enhanced Raman spectroscopy. Sharp tips, such as the points of an octahedron, help to focus the electric field.
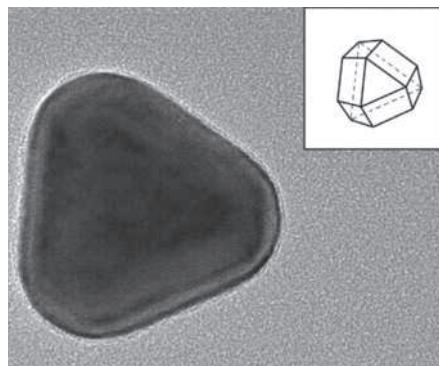
PLANT BIOLOGY
Seeding expression

Science **324,** 1447–1451; 1451–1454 (2009)

The distribution of methyl groups attached to DNA is thought to be the main route by which genes are 'imprinted', or expressed differently depending on the parent from which they are inherited. Two teams reveal that, in the plant *Arabidopsis thaliana*, extensive DNA demethylation occurs in the seed endosperm — the tissue that provides nutrients for the developing embryo — and show how it underlies imprinting.

Steven Henikoff and his colleagues at the Fred Hutchinson Cancer Research Center in Seattle, Washington, used gene-expression and endosperm-demethylation patterns to predict five new imprinted genes. And Robert Fischer, Daniel Zilberman and their colleagues at the University of California, Berkeley, found that extensive hypermethylation in the embryo accompanies endosperm gene demethylation.

The authors suggest that demethylation in the endosperm and production of small RNA molecules helps to silence disruptive transposable elements — short DNA sequences that can copy and insert themselves throughout the genome — in the embryo.

GENE REGULATION
Just-in-time activation

Genes Dev. doi:10.1101/gad.1787109 (2009)

DNA winds around bundles of proteins called histones to make nucleosomes. The histone H2A.Z — a variant of H2A — was thought to prepare genes for activation and then exit the scene once this had occurred. But Luc Gaudreau of the University of Sherbrooke in Canada and his co-workers now show that it is recruited only to the promoter regions of certain genes — those regulated by oestrogen receptor alpha — at around the time of induction. Moreover, recruitment recurs in a cyclical manner, boosting gene expression.

By identifying proteins that bind to specific DNA sequences, the researchers studied the timing of H2A.Z recruitment to a gene promoter after cells were exposed to oestradiol. Compared with H2A, H2A.Z triggers a shift in nucleosome position and stabilizes the binding of other proteins that promote gene expression.

JOURNAL CLUB

Wolf-Dietrich Hardt
ETH Zürich, Switzerland

An infection biologist points out an outstanding issue in mucosal immunology.

The gut immune system can distinguish between harmless commensal microorganisms and dangerous pathogens, and attenuates its response to the former to avoid dangerous chronic inflammation. The mechanisms that maintain this hyporesponsiveness are just beginning to be unravelled.

Dendritic cells, the key organizers of appropriate immune responses, actively sample commensal microbes. In organs other than the gut, this would trigger a strong immune response, and the responsiveness of intestinal dendritic cells to microbes is thought to be thwarted by anti-inflammatory molecules released by gut cells. But the situation could be much more complex: hyporesponsiveness might be restricted to certain 'microbe-associated molecular patterns' (MAMPs), such as lipopolysaccharides, large molecules attached to the outer membrane of many bacteria.

Linda Klavinskis of Kings College London and her team have analysed the MAMP-responsiveness of dendritic cells migrating from gut tissue to local lymph nodes. Surprisingly, these cells do respond to harmless *Bacillus* spores and most MAMPs — but not lipopolysaccharides (V. Cerovic *et al. J. Immunol.* **182,** 2405–2415; 2009). Does this suggest that hyporesponsiveness of intestinal dendritic cells is transient? The maintenance of hyporesponsiveness in the gut mucosa, patterns of MAMP-hyporesponsiveness, and localization and timing of MAMP responses will be important topics for future research.

Unfortunately, unactivated dendritic cells are hard to isolate from the gut mucosa. *In situ* analysis of dendritic-cell responses to gut microbes in intact tissue holds much promise. Technical advances in multicolour two-photon microscopy, fluorescently tagged microbes, and transgenic mice expressing cell-type and response-specific fluorescent reporter proteins will be instrumental in this key area of biology.

Discuss this paper at **http://blogs. nature.com/nature/journalclub**

NEWS

Climate talks snarled up

Two-pronged negotiations fail to bridge divide between nations.

International climate negotiators muddled through the latest round of global-warming talks in Bonn, Germany, last week, overshadowed by independent bilateral negotiations in Beijing between the United States and China.

Neither meeting produced any significant breakthroughs, and new disagreements seem to have outnumbered resolutions by a wide margin. At the United Nations climate talks that ended on 12 June in Bonn, delegates proposed so many new provisions and wording changes that the negotiation text ballooned fourfold to more than 200 pages — standard procedure for such negotiations, but one that flags up how much work remains to be done.

"We're at the point where we desperately need some higher-level leadership to get this process going," says Keya Chatterjee, deputy director for climate change at the WWF environmental group in Washington DC.

Many observers placed their hopes instead on the talks in China, which played host during 7–10 June to a US delegation led by Todd Stern, the nation's lead climate negotiator, and John Holdren, President Barack Obama's chief science adviser.

The two countries together produce some 40% of global greenhouse-gas emissions, and there is little chance of achieving an international agreement in Copenhagen, where the UN talks are scheduled to conclude in December, unless the United States and China come to

an understanding. Nonetheless, expectations going into the Beijing meeting were low.

China released a position statement on 20 May calling on developed nations to reduce emissions to at least 40% below 1990 levels by 2020. The leading US climate legislation — a bill that could come up for a vote as early as this month in the House of Representatives — would establish an emissions-trading system to reduce US emissions to 1% below 1990 levels by 2020. Other provisions in the bill would go further,

Patchy pig monitoring may hide flu threat

Public-health experts are warning that a lack of surveillance may be allowing the 2009 pandemic H1N1 flu virus to go undetected in pigs. This raises the risk that the virus could circulate freely between humans and pigs, making it more likely to reassort into a deadlier strain, they say.

Pig surveillance is largely the remit of animal-health organizations, agriculture ministries and the farming industry. Their main concern tends to be that any reports of the pandemic virus in pigs might provoke overreactions such as the mass culling of pigs that took place in Egypt, or trade bans on pigs and pork. Within minutes of the World Health Organization (WHO) announcement on 11 June that swine flu had become a pandemic, Bernard Vallat, director-general of an intergovernmental trade body, the World Organisation for Animal Health (OIE), had reiterated that trade sanctions were unjustified. "So far the role of animals has not been demonstrated in the virus's epidemiology or spread," he asserted.

But some experts say that is an artefact of patchy to

Egypt has seen mass culling of pigs (above). WHO director-general Margaret Chan (right) last week declared that the new swine flu virus had reached pandemic status.

non-existent flu surveillance in pigs. In a paper published last week (G. J. D. Smith *et al. Nature* advance online publication doi: 10.1038/nature08182; 2009), Gavin Smith, a flu geneticist at the University of Hong Kong, and his colleagues concluded that "the lack of systematic swine surveillance allowed for the undetected persistence and evolution of this potentially pandemic strain for many years". The virus originated from a mixture of swine flu strains, and pigs are an "obvious" part of the

epidemiology of the new virus, says Smith. Yet the number of swine-flu sequences in the international GenBank database is about a tenth of that for avian flu viruses. Circulation of the virus between pigs and humans is "definitely a possibility", he adds.

The pandemic virus has so far been found in pigs from just one farm, in Alberta, Canada, where it spread throughout the herd. But noone has been able to pin down how the herd became infected. Scientists at the Veterinary

Laboratories Agency in Weybridge, UK, have shown that pigs can easily become infected with the virus, and readily transmit it between themselves and shed it into the environment. Past pandemic viruses have also gone on to become endemic in pig populations.

"It's absolutely surprising that a virus this contagious in both humans and swine, and which has been reported in humans in 76 countries, has only been reported in one swine farm in Canada," says Jimmy Smith, head of livestock affairs at the World Bank in Washington DC, and a member of the organization's flu task force. "It is highly likely that more pigs are infected in more places."

Absence of evidence of the pandemic virus in pig populations is not evidence of absence, concedes Steve Edwards, chairman of the OIE-FAO Network of Expertise on Animal Influenza (OFFLU), which coordinates work done by animal-flu surveillance labs worldwide, and former chief executive of the
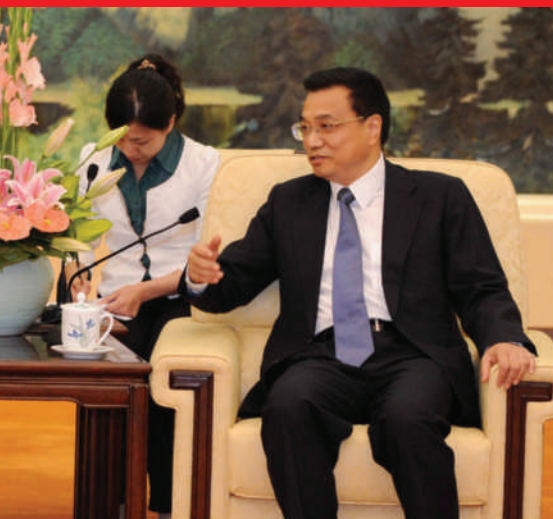
US climate envoy Todd Stern (left) in talks with China's vice-premier Li Keqiang.

RAO AIMIN/XINHUA/LANDOV

but even the most optimistic assessment, by the World Resources Institute, pegs potential reductions at only 17–23% below 1990 levels.

"We certainly did not agree with each other on everything," said Stern after returning to Washington, "but I think that we each came away with a better and a clearer understanding of each other's views and perspectives."

With just six months to go to Copenhagen and little convergence on any of the big issues, the question is whether countries are positioning themselves in expectation of striking a bargain there or merely digging trenches for a prolonged debate. Robert Stavins, an environmental economist at Harvard University, says he fears the latter might be true.

"The gulf between the countries of the industrialized world and what is usually referred to as the developing world," he says, "is, if anything, growing, or at least solidifying."

The largest pledged emissions cut on the table is of 30% from 1990 levels by 2020; that is what the European Union (EU) says it will do if others commit to similar cuts. If they don't, the EU offers only 20%. And things drop off quickly from there: Japan came under fire last week for proposing a plan to reduce domestic emissions by about 8% below 1990 levels.

Such numbers would seem to indicate a substantial divide among industrialized countries, but the gap largely disappears if the commitments are measured against a 2005 baseline.

Using this baseline, Japan's proposal is a 15% reduction by 2020; Europe's, 9–13%; and the United States', up to 10%, according to an analysis by Nigel Purvis, a former US negotiator and current president of the Climate Advisers consultancy in Washington DC. If programmes to reduce emissions internationally are included, US emissions could decrease up to 28%, with Japan's also likely to drop further.

Next up are climate discussions at the G8 meeting in July in L'Aquila, Italy, to be accompanied by a leadership session of the US-sponsored Major Economies Forum. ■

Jeff Tollefson

> "We're at the point where we desperately need some higher-level leadership to get this process going."

Veterinary Laboratories Agency. "But we should not assume it is there until we have evidence to say so," he insists.

Whereas flu surveillance has improved over the past six years in poultry and wild birds, pigs have been below the radar, says Ilaria Capua, an animal-flu expert at the Experimental Animal Health Care Institute of Venice in Legnaro, Italy. The avian H5N1 flu virus leads to serious disease in poultry and causes huge economic losses, so the OIE requires its 174 member states to report any outbreaks. In pigs, flu viruses, although common, tend to cause only mild disease, so there is no obligation to report cases of swine flu, much less take samples for genetic and antigenic analysis. The OIE has, however, asked its member states to voluntarily report any occurrences of the 2009 pandemic virus in pigs.

"Surveillance for swine flu is not something that has been high on the agenda of government services," says Edwards. "It is seen as a farming-industry problem." Most flu surveillance in pigs is passive, relying on farmers or vets sending material to government labs. Active targeted surveillance with diagnostic tests is rarer, as it is more expensive.

OFFLU has called on labs worldwide to share what information they have on swine flu, and to sequence any samples they have obtained recently. So far, however, the response has been "limited", says Edwards.

A meeting between experts from OFFLU and the WHO on 21 May — the conclusions of which were made public last week — recommended scaling up flu surveillance efforts in pigs, and OFFLU is now developing sampling and diagnostic protocols. Recommendations are all very well, but "somebody has to agree to fund all of this work", warns Edwards.

A pig of a job

Ironically, European Union funding for one of the world's largest pig surveillance networks expired in March. The European Surveillance Network for Influenza in Pigs, which was created in 2001, comprises nine European labs and one in Hong Kong. Although the network has not yet detected the new virus in pigs, its coordinator Kristien Van Reeth, an animal virologist at Ghent University in Belgium, admits that participating labs have taken just a few hundred to a thousand samples each over the past year. Network members hope that with the pandemic highlighting the need for better pig surveillance, new funding will be forthcoming.

It is now clear that the animal- and public-health communities underestimated the potential for pigs to generate a pandemic virus. Although pigs can be infected with many subtypes of flu, the three most common endemic strains are H1N1, H1N2 and H3N2. Most expected that any new pandemic would involve the introduction of a viral subtype not previously seen in humans, such as the avian H5 subtype, explains Capua. "The consensus was that a pandemic could not be caused by H1, H2 or H3 because the current human population would have antibodies against them," she says.

The emergence of the reassorted H1N1 pandemic virus — which current research indicates noone has any immunity to, apart, perhaps, from some people older than 60 — has changed that thinking. Moreover, *Nature* has learned that the international community was warned of such a risk in a presentation at a closed meeting between the OIE, the WHO and the Food and Agriculture Organization of the United Nations in Paris in February. In the presentation, the results of which are in press at the journal *PLoS Pathogens*, Capua showed that serum samples from people vaccinated against seasonal flu strains showed little or no cross reactivity against H1, H2 and H3 bird viruses, meaning that they would have no immunity.

This shows that the world needs a comprehensive surveillance system of all influenza subtypes and their evolution across many animal species, says Capua: "We should be looking at the bigger picture." ■

Declan Butler

CDC

H1N1 update

As confirmed by the World Health Organization, as of 16 June:
Cases: 35,928, in 76 countries
Deaths: 163, in 8 countries

INDIA: The health ministry announces that it will install thermal scanners in Delhi's airport to test for travellers with fever.

For *Nature*'s swine flu coverage, see
www.nature.com/swineflu.

Research parks feel the economic pinch

Once lauded as incubators of high-tech jobs, science parks find themselves struggling in the new financial environment.

Science parks, which have proliferated in recent years, face an uncertain future as the recession affects government budgets, university endowments and private investments — all of which science parks often depend on.

"A lot of big projects that were in the works are definitely feeling the crunch," says Anthony Townsend, a New York-based research director at the Institute for the Future think tank. "Everything is being delayed or slowed down right now." Further delays may be in the offing; because science parks are long-term investments, the effects of an economic slowdown may not yet be fully evident.

In the United States, many research parks are supported by state and local governments, meaning that withering state budgets can have an immediate effect. "When legislators are choosing between cutting off funding for health care for low-income children as opposed to cutting funding for science parks,

RECESSION WATCH

in most cases the children are going to win — and they probably should," says Ross DeVol, director of regional economics at the Milken Institute in Santa Monica, California.

For instance, construction has slowed on the North Carolina Research Campus in Kannapolis as park planners anxiously await a decision on state funding. The state has a projected US$4.6-billion budget shortfall for fiscal year 2010, which is likely to lead to budget cuts. Hoping that the securities markets will calm down, the city government of Kannapolis has also been waiting for more than a year to issue bonds to help fund the park.

That dependence on local governments underscores how important subsidies from universities and private companies can be for science parks. "We all need some type of subsidy — you can't just charge start-up companies the going rate," says Kevin O'Sullivan, president and chief executive of Massachusetts
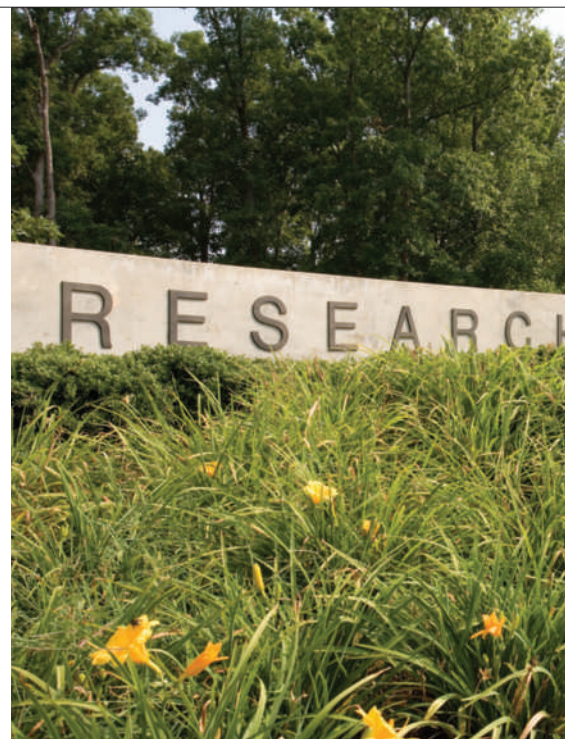
Biomedical Initiatives in Worcester, which runs three 'incubators' that foster growth in early-stage companies. Worcester Polytechnic Institute subsidizes the rent on one of O'Sullivan's buildings, and the state government has provided funds that the company uses to purchase shared equipment for its

RESEARCH TRIANGLE PARK

Astronomers lose access to military data

The US military has abruptly ended an informal arrangement that allowed scientists access to data on incoming meteors from classified surveillance satellites.

The change is a blow to the astronomers and planetary scientists who used the information to track space rocks, especially those that burn up over the oceans or in other remote locations. "These systems are extremely useful," says Peter Brown, an astronomer at the University of Western Ontario in London, Canada. "I think the scientific community benefited enormously."

When the policy changed is unclear. The website Space.com reported the end of the relationship on 10 June, but Brown says that he was told at the beginning of this year that there would be no further

data releases. Mark Boslough, a physicist at Sandia National Laboratories in Albuquerque, New Mexico, says he was told this spring that he could no longer publicly discuss the classified data to which he had some access. Neither scientist could give a reason for the end of the arrangement, and the United States Air Force, which operates the satellites, did not respond in time for *Nature*'s deadline. The Air Force did issue a 16 March memo on the military classification of fireball data, but *Nature* could not confirm its contents.

The Defense Support Program satellite network is part of the Pentagon's early-warning system. Since 1970, 23 infrared satellites in the series have been launched into geosynchronous orbit to monitor

the globe for missile launches or atmospheric nuclear blasts.

But the same infrared sensors were perfect for spotting fireballs as they streaked across the atmosphere, according to Brian Weeden, a former Air Force captain who now works at the Secure World Foundation, a non-profit organization based in Superior, Colorado. The satellites could precisely detect the time, position, altitude and brightness of meteors as they entered Earth's atmosphere. Weeden, who left the Air Force in 2007, says that the military didn't consider that information particularly useful, or classified. "It was being dropped on the floor," he says.

Under an informal arrangement, at least some of the data seem to have been provided on an ad-hoc

basis to scientists studying meteorites. Often it came in the form of an anonymous, tersely worded e-mail describing the coordinates, altitude and size of a fireball. Brown, who has collected the data since 1994, declined to specify who sent the reports.

Even the short descriptions of events were enormously helpful. In 2002, Brown and his colleagues used a larger data set from the satellites to quantify the number of objects striking Earth each year (P. Brown *et al. Nature* 420, 294–296; 2002). Last year, they were used to narrow the search for remnants of the asteroid 2008 TC$_3$ in the Sahara Desert in North Africa, and they were also crucial in recovering a meteorite fragment in 2000 from Tagish Lake in northern Canada. "In both

K. CAMPBELL/GETTY

Even the flagship Research Triangle Park in North Carolina has experienced recent lay-offs.

laboratories. But now those state subsidies, currently $700,000 a year, may be halved.

Some parks are still pushing ahead. The Bio-Research & Development Growth Park at the Donald Danforth Plant Science Center in St Louis, Missouri, opened on 16 June with more than 60% of its space already leased. Because lease negotiations for space in office parks can take a year, the St Louis space was probably filled before the economic meltdown began in earnest last autumn.

Some argue that slowing down projects can be a bad competitive decision. "It's a global marketplace now," says Gary Evans, US chief executive of Angle Technology, a consulting and venture-capital firm with US operations based in Charlottesville, Virginia. "Each project that gets delayed or shelved is just weakening the United States' competitive position overall."

In Japan, the Kansai Science City has not yet been significantly affected by the recession, says Seiji Hashimoto, planning manager of the Kansai Research Institute. He attributes this to the diversity of the park's tenants, which include research laboratories for pharmaceutical and electronics companies.

In Germany, the Innovation and Incubator Centre for Biotechnology in Martinsried has a waiting list for tenants, says Horst Domdey, managing director of Bio^M Biotech Cluster Development in Martinsried, a consulting firm that advises the park. And last autumn, the French government renewed its €1.5-billion (US$2.1-billion), three-year commitment to support the country's 71 'competitiveness clusters'.

Many are watching closely to see how science parks are faring in nations that are eager to enter the technology race. Singapore's Prime Minister Lee Hsien Loong has said that investment in its massive science-park projects, Biopolis and Fusionopolis, will continue apace despite the recession. Technology investment in the country has remained strong, as evidenced by the opening of pharmaceutical giant GlaxoSmithKline's new vaccine facility near Biopolis on 9 June. Meanwhile, Brazil, China and countries in the Middle East have continued to pour funds into burgeoning networks of science parks.

Some research parks and incubators are trying to help their tenants cope with the financial crunch. The United Kingdom's BioCity Nottingham has begun buying lab equipment and leasing it to resident laboratories. "What we don't want is a stagnating industry," says Glenn Crocker, BioCity chief executive.

For now, science parks will have to watch and wait. "We're going to get through this year, but next year are we going to have to consolidate one of our facilities? We just might," says O'Sullivan. "I'll cross that bridge when I get to it, but I'm acutely aware that this is coming." ∎

Heidi Ledford
See *Naturejobs*, page 1022.

> **"Are we going to have to consolidate one of our facilities next year?"**

---

of those cases it's hard to say whether this would have been picked up without the satellite data," Boslough says.

The data also provide a useful check against ground-based instruments monitoring low-frequency sound waves and dust from the fireball explosions, says Brown.

Although the reason for ending the arrangement remains unclear, Weeden notes that it coincides with the launch of a new generation of surveillance satellites. The US$10-billion-plus Space-Based Infrared System will provide a more detailed infrared picture of Earth for battlefield surveillance, early-warning alerts and the US missile defence programme.

The first of these satellites was launched in 2006 in a highly elliptical orbit and entered operational service in November 2008, a month after Brown

Military satellites that constitute a missile early-warning system also pick up inbound meteors.

D. SCOTT/CORBIS

received his last report. In its final configuration, the system will include at least two satellites in highly elliptical orbit and three or more geosynchronous satellites for a whole-Earth view.

Weeden speculates that the Pentagon may not want details of the new satellites' capabilities to be made public, or it may simply lack the expensive software needed to handle classified and declassified data simultaneously. "The decision may have been made that it was perhaps too difficult to disclose just these data," he says.

Brown says that whatever the reason, the end of the relationship has left the tight-knit meteorite community smarting. The global reach of the satellites and the data they supplied were unparalleled, he says.

"There's nothing else that even comes close," he says. ∎
Geoff Brumfiel

Venezuelan scientists speak out

University cuts are the latest in a series of government actions that have researchers seeing red.

Research in Venezuela, already battered on many fronts, faces budget cuts that have in recent weeks triggered students to protest and others to complain that the country's science infrastructure is approaching collapse.

Venezuela's finances depend heavily on the price of oil, which other than a recent rise has for months been below the US$60 per barrel that the government counted on when it drew up the national budget. As a result, Venezuelan President Hugo Chávez signed a decree that in March ordered government-funded institutions to reduce their 2009 spending by 6%. The cut came after five years of flat financing amid high inflation, currently close to 30%.

By law, institutions cannot save money by lowering workers' pay or by laying off staff. Because salaries make up roughly 80% of a Venezuelan university's annual spending, reductions must come from the remaining 20%. So institutions are seeing their library subscriptions and lab supplies cut back. Purse strings are likely to become even tighter in the coming months because the decree was signed after first-quarter budgets had been spent.

Yet the government claims that investment in science is growing (see page 1021). In a press conference on 3 June, science minister Jesse Chacón Escamillo said that Venezuela invests three times more in science and technology than any country in Latin America. He claimed that spending jumped from its typical level of 0.4% of gross domestic product to 1.7% in 2006 and about 2.7% in 2007.

"It is absolutely false," counters Claudio Bifano, president of the Venezuelan Academy of Physical, Mathematical, and Natural Sciences in Caracas. "It is political propaganda." On 11 June, Bifano and about a dozen university heads met to organize a study into the minister's claims. Findings are due in November.

Academic institutions can also get funding through an innovation, science and technology law, which since 2006 has required private companies to spend between 0.5% and 2% of their gross income on in-house research and development, or to give the money to an academic institution. Last year, Simón Bolívar University in Caracas received 62 million bolivars (US$29 million) this way, but how the money has been apportioned remains obscure

Cutbacks have sparked protests in Caracas against President Hugo Chávez (inset).

even to Benjamín Scharifker, the institution's chancellor. "There is a tremendous lack of accountability and transparency," he says. "We don't have any reports about the success of such grants — not even information about who are the people responsible for the projects."

Such funds are increasingly administered by government officials with military rather than science backgrounds. For example, Chacón, who declined to answer *Nature*'s questions, was a lieutenant who was involved in Chávez's 1992 failed coup attempt. "When you have people who are used to giving orders and obeying orders, it's not easy for scientists to have an open environment and to get on with our work the way we think we should do it," says Scharifker.

Not everyone is unhappy, points out Orlando Albornoz, a sociologist at the Central University of Venezuela, Caracas: "Some people believe in the government and share their views." Still, many researchers worry that academic freedom is under threat. Two months ago, scientists at both the Venezuelan Institute for Scientific Research and the Foundation Institute of Advanced Studies (IDEA) in Caracas were instructed that they could not make public statements about their work without prior written consent from the directors of their institutions. "You have no chance

of finding a position if you are considered an opponent [to Chávez] or if your name appears on a 2004 petition calling for him to go," says Luis Carbonell, president of the Venezuelan Association for the Advancement of Science's human-rights commission.

Venezuela's scientific workforce seems to have voted with its feet. In 2000, fewer than 3,000 Venezuelan scientists were registered as living in the United States; in 2008, that figure was roughly 9,000. About the same number has moved to the European Union. Some of these emigrant scientists are from the 800 researchers that the government dismissed from the Venezuelan Institute of Petroleum Research in 2003, following a national strike. Roughly 6,000 scientists remain in Venezuela.

Several lay-offs of prominent scientists have raised eyebrows recently. Jaime Requena, who had reviewed the country's scientific productivity, was fired in April from IDEA. Requena claims that he was wrongly denied a pension, wrongly accused of planning to make money from a software purchase, and that government officials told him last year they would find a way to fire him.

Like many others, Requena says that currency controls prevent international travel and collaboration. If "I am only allowed to spend $400 a year on books from abroad", he says, "how can I survive?"

Anna Petherick

J. C. SOLORZANO/MIRAFLORES PALACE/REUTERS

E. MONTILVA/REUTERS

Early man becomes early ape

An Essay on page 910 of this issue withdraws the conclusion of a 14-year-old *Nature* article (W. Huang *et al. Nature* 378, 275–278; 1995) describing what had been thought to be the oldest *Homo* fossil in Asia.

Anthropologist Russell Ciochon, of the University of Iowa in Iowa City, writes that he now thinks the jaw segment with two teeth, dated to about 1.9 million years ago from the Longgupo cave in Sichuan province, China, belongs to an ape. "We threw out a trial balloon" in 1995, he says. "Academics change their minds based on new evidence and with the passage of time."

At the time, some anthropologists had suggested that the jaw might in fact have been from an orangutan-like species (J. H. Schwartz and I. Tattersall *Nature* 381, 201–202; 1996). One of those, anthropologist Jeffrey Schwartz at the University of Pittsburgh in Pennsylvania, calls the Essay "really astonishing. It is not often that a scientist says he changes his mind. This openness is good."

The time at which *Homo* species first arrived in Asia has been a hotly contested subject; fossils of *Homo erectus* have been found dating to 1.6 million years ago in Java. The 1995 *Nature* paper implied that an earlier species, such as *Homo habilis*, was present in Asia earlier, at 1.9 million years ago.

The discovery of two apparent stone tools, the jaw and a tooth that is indisputably *Homo* — found in nearby sediments — stoked speculation that *Homo erectus* may have evolved outside of Africa. The nine authors concluded that early humans had entered Asia at roughly the same time as the genus *Homo* started to diversify in Africa. "Clearly, the first hominid to arrive in Asia was a species other than true *H. erectus*, and one that possessed a stone-based technology," they wrote.

Some anthropologists then working in China disputed this interpretation. "I never thought it was *Homo*, but akin to apes," says palaeoanthropologist Dennis Etler, now at Cabrillo College in Aptos, California. In 1998, Etler had convinced a co-author of the original paper, archaeologist Roy Larick, that the jaw wasn't *Homo*.

Ciochon says he changed his mind about four years ago after examining a tooth collection at the Guangxi Zhuang Natural History Museum in Nanning, China, including eight teeth that resemble the molars found at Longgupo. Late last year, he says, he proposed "a personal commentary" in *Nature* on the "problematic jaw".

Ciochon, the sole author of the resultant Essay, says he didn't discuss the piece with the original article's lead author and discoverer of the jaw, anthropologist Huang Wanpo of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, China, nor with co-author Gu Yumin, Huang's wife and an anthropologist. The other authors, other than Larick, weren't involved in analysing the jaw, Ciochon says.

Huang says that several scientists from around the world have long considered the jaw to be that of an ape.

Schwartz says he wants to see more details about the 'mystery ape'; Ciochon says he plans to publish those after more work with Wang Wei, director of the Nanning museum.

Rex Dalton

J. W. ROGERS
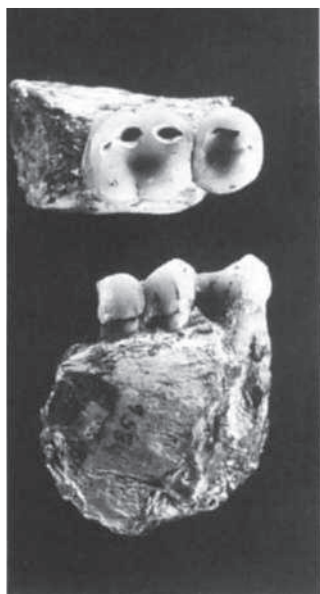
Russell Ciochon has changed his mind about the identity of fossils found in Longgupo, China.

1 cm

Italy outsources peer review to NIH

The US National Institutes of Health (NIH) is gearing up to begin a review of about 1,000 biomedical research grant applications for the Italian government, an experimental collaboration that comes at an inconvenient time for the US funding agency.

The outsourcing agreement was made last year at the request of Ferruccio Fazio, now Italy's deputy minister for health in the welfare ministry, who is looking to improve the department's peer-review system for awarding competitive research grants (see *Nature* **455,** 719; 2008). But its impact comes as the NIH deals with a flood of applications of its own, triggered by the US economic stimulus package (see *Nature* **459,** 763; 2009).

"We took on this project before the Recovery Act was passed, and we never would have taken on the Italian applications if we had known what our workload would be now," says Antonio Scarpa, director of the NIH's Center for Scientific Review. "Nonetheless, we are honoured to assist the Italians." It is the first time the NIH has provided systematic technical support for another country's grant applications, he says.

Most biomedical research funds in Italy are dispensed through government appropriations to institutions, not through a competitive grant system. Many of the national peer-review systems that do exist, and which work in response to irregular funding calls, are plagued by accusations of conflicts of interest among a small pool of reviewers.

"We want to change the culture. We need a peer-review process that is more transparent, and less prone to suspicions of bias," says Giovanni Lucignani, a diagnostic-imaging specialist at the University of Milan.

Jacopo Meldolesi, a neuroscientist at Vita-Salute San Raffaele University in Milan who ran a competitive grant programme in Italy using foreign reviewers, argues that the process might work better with Italian scientists participating in the NIH review committees.

"My reservation is that this is being done outside the Italian community," he says.

The funding at stake will provide about €29 million (US$40 million) to support 50 to 60 projects from young researchers; funds will be distributed in January 2010.

Scarpa says the NIH has agreed with the Italian government to delay the start of reviews until it has finished Recovery Act applications. "After we help coordinate reviews this year, we expect the Italians will be ready to run things on their own," he says.

Robert Paul Königs, head of scientific affairs at Germany's main funding agency, the DFG, adds that asking foreign scientists to review individual proposals is standard practice for many funding agencies. "In 2007, some 22% of reviewers consulted by the DFG were based outside Germany," he says. "But I am not aware of the evaluation of a complete set of proposals turned over to an outside agency." ■

Richard Van Noorden

Editor to quit over hoax open-access paper

The editor-in-chief of an open-access journal is to resign after claiming that its publisher, Bentham Science Publishing, accepted a hoax article without his knowledge.

Bambang Parmanto, an information scientist at the University of Pittsburgh, Pennsylvania, and editor-in-chief of *The Open Information Science Journal*, said he had not seen the computer-generated manuscript, accepted by Bentham on 3 June.

The fake paper was submitted by Philip Davis, a graduate student in communication sciences at Cornell University in Ithaca, New York, and Kent Anderson, an executive director at *The New England Journal of Medicine*. Davis says he wanted to test if the publisher would "accept a completely nonsensical manuscript if the authors were willing to pay". He retracted the paper after being notified that it had been accepted, and that he should pay US$800 to Bentham's subscription department.

Mahmood Alam, director of publications at Bentham Science Publishing, told *Nature* that "submission of fake manuscripts is a totally unethical activity and must be condemned", adding that "a rigorous peer-review process takes place for all articles that are submitted to us for publication".
For a longer version of this story, see http://tinyurl.com/lrx6m6

FDA gains the power to regulate tobacco products

L. JACK/CONTOUR

Obama in his youth.

The US Food and Drug Administration (FDA) will get the power to regulate tobacco for the first time in its 103-year history under legislation passed by Congress last week. President Barack Obama, himself a sometime smoker (see picture), has promised to sign the bill into law.

Passed by substantial majorities in both the House and the Senate, the Family Smoking Prevention and Tobacco Control Act requires that new tobacco products win pre-market approval from the FDA.

The bill bars the FDA from banning nicotine, but it gives the agency standard-setting authority that could lower nicotine levels in tobacco products. It constrains advertising, requires large warning labels on packaging and levies user fees paid by the industry to help finance FDA regulation. The fees will total $235 million in 2010.

Infrared scan reveals colourful past of the Parthenon

BRITISH MUSEUM

Conservation scientists at the British Museum in London have found the first evidence of coloured pigments on sculptures from the Acropolis in Athens. The figures formed part of the decoration on the Parthenon temple, and were taken from Greece by Lord Elgin in the early 1800s.

Ancient Greeks and Romans normally painted their sculptures, and traces of the pigments tend to survive on the objects to this day. But no hints of paint had been found on the Parthenon sculptures despite detailed studies — including an analysis in the 1830s by English physicist Michael Faraday.

The researchers revealed the presence of a pigment known as Egyptian blue on the belt of the goddess Iris (pictured). They used a portable detector to beam red light onto the surface and capture the infrared light emitted by the luminescent pigment particles (inset).
For a longer version of this story, see http://tinyurl.com/m24ylw

Japan's lunar orbiter ends mission with crash landing

The Japanese space agency's KAGUYA lunar orbiter ended its 21-month mission with a planned crash into the Moon on 10 June.

Formally known as SELENE (Selenological and Engineering Explorer), the mission was launched in September 2007. The orbiter gathered detailed geological information about the Moon, mapping its gravitational field and taking high-definition video images.

As *Nature* went to press, NASA's Lunar Reconnaissance Orbiter was scheduled to launch on 18 June. Together with India's Chandrayaan-1 spacecraft, which launched in October 2008, it will attempt to spot water ice at the Moon's poles (see *Nature* 459, 758–759; 2009).

Artefact raiders charged after undercover operation

A two-year federal investigation of widespread Native American grave robbing and artefact theft in the Four Corners region of the United States culminated in charges against 24 people last week.

The arrested individuals, from Utah, Colorado and New Mexico, were arraigned in the federal court on 10 and 11 June for multiple felony indictments for trade in 256 artefacts with a total value of more than US$335,000. Purloined items included pottery, baskets, sandals and necklaces taken from excavations on federal lands.

An undercover agent purchased the looted artefacts from, among others, a high-school teacher and an honoured archaeological-tourism promoter. One man — James Redd, a physician from Blanding, Utah — committed suicide on 11 June, authorities say, the day after he and his wife were charged with artefact theft.

US revives FutureGen 'clean' coal plant

The US Department of Energy (DOE) has announced plans to revive FutureGen, a commercial-scale coal-fired power plant in Mattoon, Illinois, that would capture carbon dioxide emissions and sequester them underground.

Under George W. Bush's administration, the DOE pulled the plug on the flagship 'clean' coal technology programme in January 2008, citing a dispute with industry partners over the US$1.8-billion price tag (see *Nature* 451, 612–613; 2008).

Energy secretary Steven Chu revived FutureGen on 12 June, announcing an agreement to restart the negotiations, update the cost estimate and begin preliminary design activities. The DOE and FutureGen industry partners hope to make a final decision on whether to go ahead with the project early in 2010.

Corrections
The News Feature 'Sucking it up' (*Nature* **458,** 1094–1097; 2009) incorrectly stated that Global Thermostat is waiting for venture-capital funding to build a prototype for the capture of CO_2 from the air. It already has sufficient funding in place.

The News story 'Funding struggle for mercury monitoring' (*Nature* **459,** 620–621; 2009) erroneously located Changbai Mountain in Taiwan. It is in northeastern China.

In the News Feature 'Nascence man' (*Nature* **459,** 316–319; 2009), the picture of the Lost City on page 318 should have been credited to D. S. Kelley.

The NEW & IMPROVED KELVIN

Metrologists are on a path to redefine the unit of temperature. The freezing point of water will never be the same again, finds **Nicola Jones**.

A little over 20 years ago, Mike Moldover played a tune through a sphere of argon gas and became something of a rock star in the exacting field of metrology. The way those notes reverberated enabled him and his colleagues to determine the value of the Boltzmann constant — the relationship between the average kinetic energy of molecules and temperature — better than anyone ever before. Moldover was sufficiently confident of the result to put a bold promise into his paper: "if by any chance our value is shown to be in error by more than 10 parts in 10^6, we are prepared to eat the apparatus" (M. R. Moldover *et al. J. Res. Natl Bur. Stand.* **93,** 85–144; 1988). No one expects Moldover to have to swallow his argon anytime soon.

Moldover, who worked at the US National Institute of Standards and Technology in Gaithersburg, Maryland, and his colleagues determined the value of the Boltzmann constant (k_B) as being $1.3806513 \times 10^{-23}$ joules per kelvin with an uncertainty of just 1.8 parts per million (p.p.m.). A generation later, that result continues to inspire others. "Scientists are sort of in awe of Mike," says Michael de Podesta of the National Physical Laboratory in Teddington, UK. Today, de Podesta is racing others around the world to whittle down the error in Moldover's experiment to just 1 p.p.m. and at the same time confirm his result. Once the constant has been measured to that level of accuracy, an international committee plans to set it in stone, and use it to fundamentally shift the definition of the unit of temperature. "It's the hope that once this is settled, the world will stop worrying about thermometry," says Moldover.

The worry does not, admittedly, keep many people awake at night, nor is it of immediate practical importance. But it does have a certain philosophical heft. Today, the SI unit of thermodynamic temperature, the kelvin, is defined in terms of absolute zero and the triple point of water (the temperature and pressure at which water exists as a solid, liquid and gas in equilibrium), which is fixed at 273.16 K. This was done for historical reasons that were logical at the time. But to metrologists, the definition is inelegant, illogical and downright irritating.

"It's a slightly bonkers way to do it," says de Podesta. According to the metrological code of ethics, it is bad form to grant special status to any single physical object, such as water. Worse, 'water' needs to be qualified further: at present, it is defined as 'Vienna standard mean ocean water', a recipe that prescribes the fractions of hydrogen and oxygen isotopes to at least seven decimal places. Finally, it makes no sense for researchers studying the chilly climes of low-temperature physics or the blazing heat of stars to have their measurements of

M. BREGA

temperature rely, even theoretically, on comparison to a random piece of ice. "The kelvin is ripe for redefining," says Richard Davis, one of Moldover's former colleagues, who works at the International Bureau of Weights and Measures (BIPM) in Sèvres, France.

The BIPM, which is in charge of such things as units, aims to fix this unconscionable situation by 2011. The Boltzmann constant, which is currently a measured value with some uncertainty, will be declared a set number (letting it live up to its name of 'constant') and the kelvin will be redefined as something like the change in thermodynamic temperature that results in a change of mean translational kinetic energy of $1.38065XX \times 10^{-23}$ joules — the current work will fill in the Xs.

Sound science

The change represents a metrological sleight of hand. Whatever uncertainty existed in the measure of the Boltzmann constant will be wiped clean by definition, and transferred to uncertainty in the measure of temperature. These new temperature uncertainties are so tiny that they certainly won't cause panic in any lab — most researchers won't even notice them. Instead the redefinition will open the door for improved accuracy at the far ends of the temperature spectrum. More than that, it frees the kelvin from an unsteady connection to water and rests it on the intellectually firmer foundations of a physical constant. As such, the move mirrors what metrologists are doing with all the units of measurement.

In the case of the kelvin, the technique of choice for determining k_B is called acoustic thermometry, and it relies on being able to precisely determine the speed of sound in a gas-filled sphere at a fixed temperature. From this the Boltzmann constant falls out of the equations relating the kinetic energy (E) to the thermal energy of the gas ($E = 1/2mv^2 = 3/2k_BT$, in which m is the mass of one atom, v is the average speed of the atoms — which is proportional to the speed of sound in the gas — and T is the temperature). The speed of sound can be measured by analysing the frequency of the sound waves that resonate within the sphere. The good news is that experimentalists can measure the frequency with great precision; de Podesta has a rubidium clock in his lab that's accurate to 1 part in 10^{13}. The bad news is that to analyse the resonances accurately, the volume of the container must be known with great precision. Moldover did this in his groundbreaking 1988 experiment by first filling his 3-litre sphere

with mercury and weighing it. (He promised in his paper to eat this, too, should his result be proved wrong).

The volume issue remains the hardest problem to crack. Filling the spheres with a liquid and weighing them is still a possible solution. But there is a more elegant method, the details of which have just been worked out by one of Moldover's original colleagues, James Mehl, professor emeritus at the University of Delaware in Newark. The idea is to fit the sphere with tiny antennas to measure the resonance of microwaves in the cavity, then use this measurement to determine the sphere's volume. The researchers will deliberately use a lopsided sphere, such as one with three axes of slightly different lengths, to avoid the complicated pattern of overlapping resonances that results from using a sphere approaching, but not reaching, perfection.

"The kelvin is ripe for redefining."
— Richard Davis

"It needs to be very exactly not a sphere," says de Podesta. To reach that kind of precision, he uses the same manufacturers that are involved in producing the mirrors for the future James Webb Space Telescope.

The effort to best Moldover has inspired a friendly race between three main laboratories: the National Physical Laboratory, the National Institute of Metrological Research in Turin, Italy, and France's equivalent national lab outside Paris. All three are using acoustic thermometry — although with slightly different apparatus — to determine k_B. Every aspect

of their work requires extreme care. "You buy a very pure gas, which isn't pure enough. So you purify it more," says Roberto Gavioso from the institute in Italy, explaining just one technical detail from a long, long list. One or all of the labs should hit 1 p.p.m. by the end of 2009 or 2010. "I'm cheering for these guys," says Moldover who, now 69, has passed on the Boltzmann torch.

At the same time, the German National Metrology Institute in Berlin, heads an alternative approach. Bernd Fellmuth's kit consists of a pressure chamber surrounded by a cubic metre of water to keep the temperature constant. A set of capacitors inside the chamber measures the capacitance before and after helium is inserted, which allows the team to calculate k_B using a different set of equations from those used by the acoustic thermometry group. A difficulty here, however, is how to measure the pressure to the required accuracy; currently that's only possible to about 4 p.p.m., and an entire team is required to reach that point. The group hopes to begin experiments in January 2010 that could get the accuracy down to 1 or 2 p.p.m. by the end of that year. Moldover calls that an optimistic target, but it would be good to have a check for the acoustic work in time for 2011, he says.

Once the results are in, the decision goes to an international vote. Each of the 53 nations that are members of the BIPM supplies delegates to the General Conference on Weights and Measures, which will make the final decision. This high-level conference meets once every four years — the next being in 2011.

Why the rush to this finish line for the kelvin?
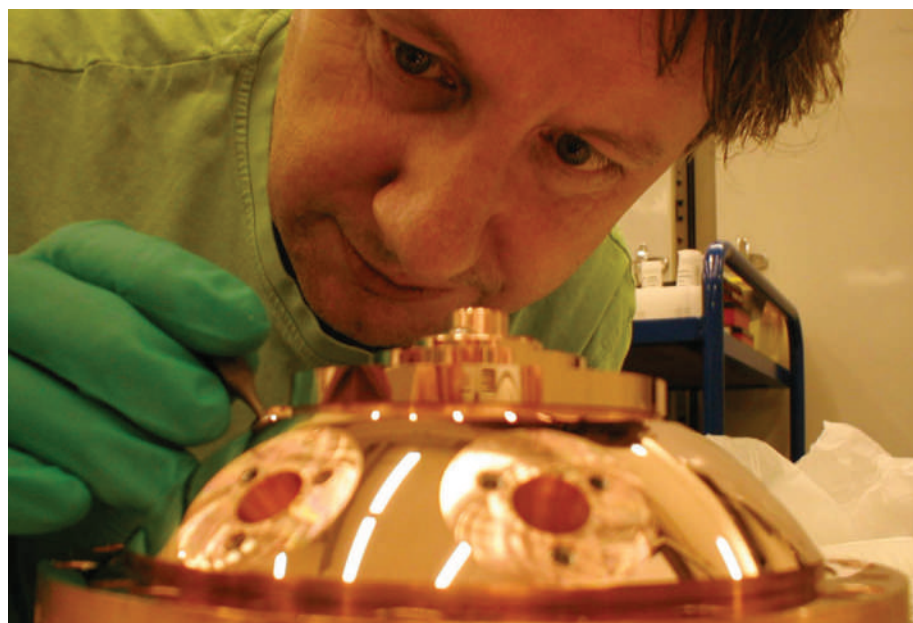
A. HEUME

Michael de Podesta and his team is trying to lessen the error margin in measuring the Boltzmann constant.

A lump of metal serves as the kilogram standard.

It is all the kilogram's fault. The kilogram is the last remaining quantity that is defined by a single, physical object — a lump of platinum and iridium, dubbed Le Grand K, that is held in a BIPM vault on the outskirts of Paris. This has the curious effect that the value of one kilogram, although meant to be constant, almost certainly changes as atoms are added to or brushed off the surface of Le Grand K. No one knows exactly how unstable this value is, of course, as there is no fundamental standard with which to compare it. This is hardly firm bedrock on which to rest all measurements of mass.

To solve this problem, various research teams have been seeking to redefine the kilogram on the basis of a universal quality. The most intuitive technique involves making a perfect sphere of silicon-28, such that the number of atoms within the sphere can be precisely determined. Thus a kilogram could be redefined as the mass of a set number of silicon atoms, although in practice this would be almost impossible to achieve because making a perfect enough sphere is tremendously difficult to do.

A more practical method involves a 'watt balance', which measures the electromagnetic force required to counterbalance a kilogram under Earth's gravity. The equations describing that electrical force involve Planck's constant, a fundamental parameter of quantum physics. Right now, like the Boltzmann constant, Planck's constant must be measured, but if it were given a set value, the kilogram could be redefined on the basis of that number. To accomplish this feat, metrologists are trying to determine Planck's constant with an uncertainty of less than the instability thought to be associated with Le Grand K. That outdated object, having been used as a standard to set the value of the Planck constant, could then be retired.

Both methods are being actively pursued, as checks to one another, and researchers are keen to change the definition of a kilogram as soon as possible as it will stop the 'drift' in its value.

As work on the kilogram progresses, metrologists have resolved to tie up all their other loose ends. In 2005, they decided to reconsider the kelvin, the kilogram, the ampere and the mole — four of the seven 'base quantities' — the other three being the metre, the second and the candela. These four all have definitional deficiencies and are ripe for updating. The redefinition of the ampere and the mole is linked to the work on the kilogram. But the kelvin stands alone.

So thermometry specialists press on with their efforts to tidy up their corner of the metrological universe, even though this work will make little immediate difference to the world. To keep science and industry on an even keel, nothing will change initially except the definition (see 'A temperature for the masses').

That lack of impact might provoke existential stress for some scientists. "Sometimes I feel a bit guilty because maybe I should be helping starving children or something," says de Podesta. But he says he feels deeply that it really matters that there are people worrying about tiny fractions of a kelvin and about laying a foundation of truth in thermometry. One day, someone will need accurate and precise thermodynamic temperature readings, and having a fixed value for k_B will help them. "I'm doing a small thing," says de Podesta. "But I do think it's a good thing for the world." ∎

Nicola Jones is a commissioning editor for *Nature*'s Opinion section.
See Editorial, page 890.

A TEMPERATURE FOR THE MASSES

Efforts to redefine the unit of thermodynamic temperature by 2011 will not affect most scientists because they deal with something subtly different — a temperature scale called T_{90}.

The thermodynamic temperature of a substance is the average kinetic energy of its atoms. To determine this accurately and directly requires exacting techniques, such as measuring the reverberations of sound in a gas (see main text). This is regarded as primary thermometry. But most labs measure temperature with simpler devices, such as a platinum wire, the resistance of which changes at different temperatures. These thermometers can be very reproducible (often within the startlingly precise margin of 0.1 parts per million) if calibrated at a national metrological lab. There, scientists measure the wire's resistance at several set phase transitions, then develop an equation relating resistance across the wire to temperature.

To do this, the national labs need to know the thermodynamic temperatures of the key phase transitions. These values have been measured many times, using a plethora of primary techniques that often do not agree precisely. So to ensure that all national labs are on the same page, the International Bureau of Weights and Measures in Sèvres (BIPM), France, compiled a book of reference transition temperatures in 1990 called the ITS-90 (for 'international temperature scale'). Thermometers calibrated against this, including common lab thermometers, give readings on the T_{90} scale.

But many of the temperatures listed in the ITS-90 are now known to be wrong. Advances in primary thermometry have shown that some of these are based on flawed measurements. The freezing point of gold, for example, is listed in the ITS-90 as 1,337.33 K, and that's reproducible between labs to within 0.01 K. But more accurate primary thermometry reveals that the true freezing point is closer to 1,337.38 K.

Although the reproducibility keeps engineers happy, the inaccuracy annoys metrologists. Metrologists dream of a day when a primary thermometer is invented that is as easy to use and as reproducible as a platinum wire — and as accurate as today's best devices. Then, metrologists could kill the ITS. In this future scenario, the uncertainty in temperature measurements would be minimized if the Boltzmann constant (k_B) — a value that relates kinetic energy to temperature — were fixed, as is planned through the redefinition of the kelvin.

The ITS-90 errors are greatest at the extremes of temperature — below about 20 K or above 1,300 K — and scientists working in these regimes often use primary thermometry, so fixing k_B makes sense for them. It would, for example, fit well with the way that astronomers work, as they can't stick a thermometer into a star. But errors of 1 part per thousand are sufficient for many astronomers, so improving the accuracy to 1 part per million is unlikely to make any practical difference for them now. And some people aren't even aware of the intention to redefine the kelvin: Horst Meyer, co-editor of the *Journal of Low Temperature Physics*, didn't know about it until told by *Nature*.

"That's the whole point; you shouldn't notice," says Richard Davis, executive secretary of the BIPM committee that is working out the details of the new definition. "Things are wrong if it gets to a state that you notice." **N.J.**

Planting the forest of the future

While conservation biologists debate whether to move organisms threatened by the warming climate, one forester in British Columbia is already doing it. **Emma Marris** reports.

At a research station in the Okanagan valley in British Columbia, a few kilometres outside the town of Vernon, orderly rows of trees run alongside the road. Many of the conifers stand tall and full, producing seeds destined for plantations around the province. But one dusty brown field is filled with lines of seedlings just ankle high. Greg O'Neill, 45, who planted the trees in April, walks among them with a slightly paternal air. They are part of a very slow experiment that will yield the first results that are useful to policy-makers just about in time for him to retire.

In a world in which many scientists fret about the toll that global warming is exacting on nature, O'Neill is actually doing something about it. A research scientist for the British Columbia Ministry of Forests, he is moving seedlings to areas that are outside their current comfort zone to test how they might handle the warmer conditions of the future. And he is behind a government push to move tree populations into new areas to prepare them for the warming climate.

Some hail the policy as pragmatic and forward-thinking, but others label it as dangerous

> "Douglas fir grows from Mexico to British Columbia, but move it 700-metres elevation downhill and you will be growing toothpicks."
> — Greg O'Neill

and premature. Plants moved by humans may become invasive in their new haunts or just fail to thrive. Yet O'Neill thinks the potential gains in terms of timber production and forest health are worth the risk.

The Canadian province is a pioneer in the field of adaptation, according to Gerald Rehfeldt, a retired United States Forest Service geneticist based in Moscow, Idaho, who has long championed moving trees. "British Columbia is ahead of everybody in the western hemisphere," he says.

O'Neill's experiment is called the Assisted Migration Adaptation Trial (AMAT) and it is what foresters call a provenance trial. The AMAT is taking seedlings from 40 spots in British Columbia, Washington state, Oregon and Idaho and planting them in 48 sites all over those same areas. All told, 16 species are involved in the project (see 'Testing assisted migration').

One element of the project tests whether moving trees north will enable them to fare better as the climate changes around them. Another part of the experiment turns the clock forwards on the forests by taking commercially

important trees and moving them south, forcing them to endure a warmer climate, quickly simulating years of climate change.

O'Neill expects 50% of his seedlings to die, mainly because trees are highly diverse genetically and are often adapted to specific local climates. Take a Douglas fir from the coast and bring it inland, and it may well die within a few years. If it doesn't, the chances are that it will be small, pest ridden or crooked. "Douglas fir grows from Mexico City to central British Columbia, but move it 700-metres elevation downhill at any location, and you will be growing toothpicks," says O'Neill.

Warmer woods

A prototype spruce-only provenance trial on an adjacent patch gives a preview of how some of these seedlings might look in five years. A few are about chest height, others are half or a quarter as tall. Some are dead or half dead, their tops dry and bent into a 'shepherd's crook', invaded by white pine weevil. In general, ones from farther away fare worse.

O'Neill sees a similar pattern at older provenance plantings, some now 50 years old. These trials were planted before climate change was an acknowledged problem. The idea was to see if new combinations of populations and areas could lead to better yields — this works

Greg O'Neill surveys a 5-year-old experimental planting in British Columbia.

because trees are not always particularly well adapted to their locations.

Throughout the province, O'Neill thinks that many trees are already feeling the effects of climate change. "We see some really ugly trees out there," says O'Neill. "Maybe the pest and pathogen outbreaks we are seeing are compounded by that maladaptation."

The mountain pine beetle has spread through the province in recent years, turning healthy lodgepole pines a desiccated red colour as it kills them. So far, 14.5 million hectares are affected[1] — an area larger than Greece. The warming of the climate is widely thought to be a factor in the outbreak — usually low temperatures in the winter would kill off much of the beetle population.

For the Northern Hemisphere, climate models agree that regions in the north will warm more and faster than areas nearer the equator. British Columbia had already warmed 0.7 °C in the decade to 2006 (ref. 2) — almost as much as the globe has warmed over the past century — and a middle-of-the-road climate simulation for British Columbia projects another half a degree increase every decade. The province is not alone in facing changes to its forests. One analysis suggests that 88% of the western United States will see a turnover in the kinds of tree communities it hosts by the end of the

century[3]. On the other hand, the Intergovernmental Panel on Climate Change predicts that rising carbon dioxide levels will lead to more growth, creating bigger trees and boosting the forestry industry in some regions[4].

If foresters stopped replanting trees, the wind and birds would carry pollen and seeds in all directions, and trees would thrive wherever seeds landed in a spot with a suitable climate. But judging from the results of past provenance tests, says O'Neill, this would be no slow and stately march. It would be more like a scrappy scramble, with few trees doing very well, and individual types moving in different directions at different speeds. For a long time to come, forests would probably look a bit more raggedy, stunted and jumbled. Not ideal for producing timber.

The possible changes were brought home to foresters in British Columbia by a 2006 study[2] containing maps of what the climate in British Columbia might look like in 2025, 2055 and 2085. It shows that forestry zones — which include trees and their environment — will shift substantially. The 'climate envelope' of trees whose ranges now end in the province will move at least 100 kilometres north a decade. For example, a ponderosa-pine ecosystem, now found only in sunny spots around the Vancouver area, may well spread into the boreal forest in the northeast of the province. But dozens of factors, including the trees' genetic variability, the direct effect of increased carbon dioxide on growth and the complexities of migration will affect whether the trees make it all 100 kilometres.

For British Columbia, forecasts of natural tree movement tell only part of the story because most new trees there are planted by

Examples of tree species being planted in assisted-migration plots.

people. Of the province's 60 million hectares of forest, 25 million hectares are commercial forests, and when one area is logged, trees must be replanted. To foresters, the projections of future change looked a bit like an instruction manual of what to plant where. "When they showed that the climates represented by the forestry zones would shift really huge distances, it really hit home much more than saying the climate would warm 4 °C," says O'Neill.

Layers of fear

The forest forecasts were just one of the things that scared foresters and policy-makers in British Columbia into action, says O'Neill. Another was the mountain pine beetles. "People see red trees as far as the eye can see and it becomes easier to implement some rather aggressive policies," he says. The final things that crystallized opinion were the forest fires caused by drought in 1998 and 2003.

As a result, in April, the province started systematically moving seeds uphill in what ecologists call the first widespread instance of assisted migration. The action came about through a change in the rules governing how timber companies replant. British Columbia, like most other wood-producing areas, has planting policies designed to take advantage of the local adaptation of trees. After companies log an area, they must chose seeds from nearby for replanting. Rules vary by species but, before this year, generally, a seed could be moved only within about 200 kilometres north, south, east or west and 200 metres downhill or up to 300 metres uphill from its origin.

The new rule stretches the elevation guidelines so that seeds can be moved up to 500 metres uphill, in accordance with the idea that moving trees substantially higher in elevation will help them thrive in conditions to come. "These changes encourage planting seed that is adapted to climates about 2 °C warmer than

TESTING ASSISTED MIGRATION

M. CARLSON

N. UKRAINETZ
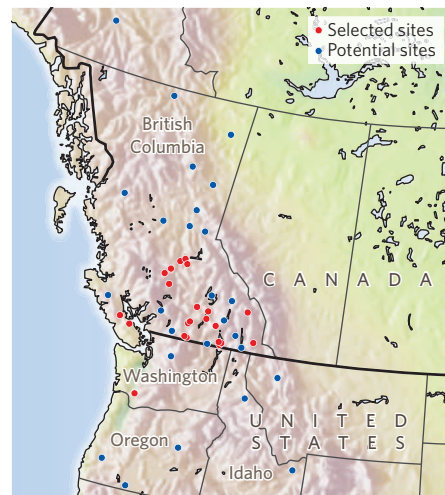
Whitebark pine is threatened by global warming.

J. WILBURN/PHOTOLIBRARY.COM

the planting site, to account for climate change in the past century and changes anticipated in the first portion of the planted tree's lifetime," says O'Neill.

According to Jim Snetsinger, British Columbia's chief forester, foresters are fans of the change and eager for even more flexibility in policies so that they can chase the climate further. "They're looking for something bigger," he says.

Another potential action would be moving species outside their historical ranges. Neither O'Neill in his AMAT nor forestry companies have gone that far; they are moving populations of trees around within their historical range. But O'Neill says the more substantial movement is a natural extension of what he is doing.

Upping the stakes

So far, the tree transplanting has happened only with commercial timber. But O'Neill is talking about "taking it to a new level" by putting in the odd hectare of southern seed in places such as parks and extremely remote forests in the province, places not likely to be logged and replanted for a long time, if ever. He's calling the idea "directed diaspora". When the temperature has risen, O'Neill hopes that the genes to adapt trees to the heat will already be circulating. "Why not give them a boost?" he says.

But many conservation biologists think that moving species around is a bad idea. Humans have carted plants and animals all over the globe, intentionally and unintentionally, for millennia. Some of these movements have gone spectacularly badly when the 'invasive' species have been successful enough to drive native species to extinction, especially on islands and in lakes. Invasive pine species have swamped the diverse fynbos shrublands of South Africa, sucking up the area's water in the process, for example.

The Monterey pine has been on both sides. Its native range includes a tiny sliver of the California coast and a couple of islands off Baja Cal-

E. MARRIS

ifornia in Mexico, where it is considered to be threatened by development and by an invasive tree disease. But the pine has itself turned into an invader in Australia, South Africa and several other Southern Hemisphere countries, where it was planted for timber.

Daniel Simberloff, an ecologist at the University of Tennessee in Knoxville specializing in invasion biology, calls British Columbia's programme "a waste of time". He sees the whole notion of assisted migration as a bandwagon that has lately become chic. He is less worried about moving populations within the historical range of a species than with moving a species beyond its range. But O'Neill's AMAT trial won't be able to inform policy for a long time, says Simberloff. "He'll be dead before there are any real data from this." Meanwhile, Simberloff says, there is just too great a chance that the translocated trees, or the diseases they host, will become invasive. "I would want to know a lot more about pathogens and insects before I moved things," he says. "There is very little evidence that it is going to help, and in addition there are potential downsides."

So far, critics have either not noticed policy changes in British Columbia or felt that they have yet to cross a line. "We haven't heard anything negative from anyone," says Snetsinger.

Moving trees is also being mooted in Europe. But, Csaba Mátyás, a forestry researcher at the University of West Hungary in Sopron, says he hasn't seen scientific concern spill over into action yet. Nevertheless, he and his European colleagues are pressing for change. And he suggests that the European public won't be very worried about moving populations or spe-

cies around. "This is not an issue in Europe," he says, "as this has been done by foresters for centuries."

Although conservationists are often opposed to moving species, some have started to reconsider assisted migration (also called managed relocation and assisted colonization). Some conservationists see it as the only way to save some slow-moving species from climate change. The idea has been proposed for the California quino checkerspot butterfly and the whitebark pine, whose seeds are an important food for some bear species. But among even the most vocal proponents of this tactic, some think that O'Neill is going too fast.

"This experiment is wonderful, I am glad he is doing it," says Dov Sax, an ecologist at Brown University in Providence, Rhode Island. But, he says, the idea of moving seeds into natural forests gives him the shivers. "The thought of planting a couple of hectares of trees far to the north of where people currently harvest — that sounds a little scary to me." Sax calls for more study, and no precipitous moves. But he admits that his call for caution is as predictable as O'Neill's eagerness to move ahead. "You can see why people with a commercial interest are going to want to find a solution as quickly as possible. Conservationists, who have seen experiments go awry, are going to be more cautious."

> "I would want to know a lot more about pathogens and insects before I moved things."
> — Daniel Simberloff

The nervousness of conservation biologists and ecologists can also be partly explained by the fact that their subjects and, often, their passion, are ecosystems and all the co-evolved interactions that compose them. Although they want to save species from extinction, ideally they want to save species by saving ecosystems. When trees from farther south suddenly appear, the ecosystem has changed, even if no species are harmed.

In the end, if O'Neill decides to move trees as he would like, it won't matter much what conservationists decide. Even if his directed diasporas are banned from some tracts of land, those protected areas will still be assailed by the spread of seeds and a haze of lime green pollen drifting from commercial plantings. ∎

Emma Marris writes for *Nature* from Columbia, Missouri.

1. www.for.gov.bc.ca/hfp/mountain_pine_beetle/facts.htm
2. Hamann, A. & Wang, T. *Ecology* **87,** 2773–2786 (2006).
3. Rehfeldt, G. *et al. Int. J. Plant Sci.* **167,** 1123–1150 (2006).
4. Easterling, W. E. *et al.* Food, fibre and forest products. in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 273–313 (Cambridge University Press, 2007).

A seedling planted this year in British Columbia.

CORRESPONDENCE

Global warming: why the 2°C goal is a political delusion

SIR — The papers by Malte Meinshausen and colleagues ('Greenhouse-gas emission targets for limiting global warming to 2°C' *Nature* **458,** 1158–1162; 2009) and by Myles Allen and colleagues ('Warming caused by cumulative carbon emissions towards the trillionth tonne' *Nature* **458,** 1163–1166; 2009) suggest that society could limit global warming below the widely discussed goal of 2°C by adopting a cumulative budget for carbon emissions. Although they do underscore the difficulties, their prescriptions are only marginally relevant for policy design.

Solving the carbon problem needs international coordination. Success depends on many factors, but paramount is the credibility of promises that governments make to each other through international agreements. The trouble with the Kyoto treaty was that for pivotal countries, notably the United States, the promises were not credible. Correcting that error is a central aspect of negotiations before the climate summit in Copenhagen in December.

Credible promises will make most countries willing to do even more: a cycle of cooperation could unfold. Essentially, all successful international regulatory regimes evolve this way, starting with modest promises that, if kept, create confidence and credibility for greater efforts later on.

The problem with long-term cumulative targets such as those Allen advocates is that they cannot readily be codified into anything governments will find credible. They lack immediacy for policy if governments decide to leave costly actions to their successors.

This is partly why Kyoto's 'budgets' lasted only five years (2008–12). Nobody thought that was long enough, but it was expected to force action to smoke out credibility. (In the United States, alas, the effort failed.)

Global, cumulative emission budgets are nothing new. But they will never gain traction because a government must translate them into something it can control, such as shorter-term emission targets; it can implement these through 'cap-and-trade' schemes or other kinds of tangible policy effort, such as carbon taxes or regulatory programmes. At best, broad cumulative budgets are a general guide for policy. At worst, they distract the debate from what governments can actually achieve.

Your special issue of 30 April 2009 on 'The coming climate crunch' is also a timely reminder that the 2°C target is a political delusion. Nobody knows what is safe — in part because the climate will be sensitive in unknown ways (as Meinshausen's paper shows) and also because safety depends on circumstances. There is no simple relationship between what governments can actually control and abstract goals such as a set limit to warming. Real outcomes might be plagued by interactions that doom the planet to warming of 2°C (or more), whether or not emissions are cut. Even with a big dose of luck, the effort needed to get to 2°C would be heroic, as Allen and colleagues indicate, and probably far beyond what real governments can achieve.

Being neither achievable nor safe, the target is becoming dangerous. The new papers are a reminder of how wrong-headed such goal-setting has become.
David G. Victor Energy and Sustainable Development, Stanford University, Stanford, California 94305, USA
e-mail: dgvictor@stanford.edu

Ancient ivory figurine deserves a more thoughtful label

SIR — In his News & Views article 'Origins of the female image' (*Nature* **459,** 176–177; 2009), Paul Mellars describes the 35,000-year-old figurine of a woman, carved from a piece of mammoth ivory, as "explicitly

— and blatantly — that of a woman, with an exaggeration of sexual characteristics (large, projecting breasts, a greatly enlarged and explicit vulva, and bloated belly and thighs) that by twenty-first century standards could be seen as bordering on the pornographic".

Mellars has, of course, never been pregnant. Anyone who has would know that breasts of that size (given the unavailability of surgical intervention at the time) are evident only in the late stages of pregnancy and during lactation. Likewise, it seems a stretch to imagine that a woman who was eking out an existence many millennia ago would be carrying around so much extra body fat — unless her "bloated" belly and thighs were the result of a pregnancy. Also, a "greatly enlarged" vulva is one of the more obvious ramifications of an infant making its way through a passage narrower than its head.

For this reader, the figurine speaks across the ages of fertility, not sexuality, and certainly not of pornography. It could have been carved as a pendant in the hope that it would provide its wearer with a talismanic connection to the power and mystery of creation — and not, as media headlines have described it, as a piece of "prehistoric porn".

It is unfortunate, then, that the figure accompanying Mellars's piece is captioned "A 35,000-year-old sex object". By the time it appeared in *Nature*'s video archive, its title had become the rather more risqué "Prehistoric pin-up". And when the story hit the Internet, this groundbreaking discovery of the oldest piece of figurative art known to humankind was labelled "'Porn' art in ivory, 35,000 years old".

This misguided focus on a salacious interpretation has caused a cascade effect that trivializes and coarsens a monumental scientific and artistic discovery.
Anna McDonnell 2436 3rd Street #B, Santa Monica, California 90405, USA
e-mail: annamcdonnell@mac.com

San Andreas array failure is only a temporary setback

SIR — In your News story 'Geologists suffer observatory glitches' (*Nature* **459,** 20–21; 2009), the presenter at a recent meeting (M.D.Z.) is described as having "let slip an embarrassing fact" — that an array of instrumentation deployed within the San Andreas fault zone at a depth of 2.6 kilometres became inoperable soon after installation. We disagree with this description of what was said. The status of the instrumentation was one of the principal topics of the presentation. The agencies funding the San Andreas Fault Observatory at Depth (SAFOD) — the National Science Foundation and the US Geological Survey — have known about the status of the array since it stopped working seven months ago.

The premature failure of the array deployed last September was a disappointing, but only temporary, setback. Because such instruments have never operated at depth at high temperatures and pressures for long periods of time, the array was designed to be retrieved and refurbished every few years. We shall therefore be able to recover the equipment and diagnose and correct the problems that caused it to cease operation.

The acquisition of unique data and samples from depth within the San Andreas fault zone will allow long-standing fundamental questions about earthquake processes to be addressed. We, and the hundreds of scientists and engineers from around the world who contributed to the success of SAFOD, are proud of these accomplishments.
Mark D. Zoback Department of Geophysics, Stanford University, Stanford, California 94305, USA
e-mail: zoback@stanford.edu
William Ellsworth, Stephen Hickman US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

ESSAY

The mystery ape of Pleistocene Asia

Fossil finds of early humans in southeast Asia may actually be the remains of an unknown ape.
Russell Ciochon says that many palaeoanthropologists — including himself — have been mistaken.

Fourteen years ago, a *Nature* paper by my colleagues and I described a 1.9-million-year-old human jaw fragment from Longgupo in Sichuan province, China[1]. The ancient date in itself was spectacular. Previous evidence had suggested that human ancestors arrived in east Asia from Africa about 1 million years ago, in the form of *Homo erectus*. Longgupo nearly doubled that estimate. But even more exciting — and contentious — was our claim that the jaw was related to *H. habilis*, a species of distinctly African origin. If this descendant of *H. habilis* had arrived so early into southeast Asia, then it probably gave rise to *H. erectus* in the Far East, rather than *H. erectus* itself sweeping west to east.

For many years, I used Longgupo to promote this pre-*erectus* origin for *H. erectus* finds in Asia. But now, in light of new evidence from across southeast Asia and after a decade of my own field research in Java, I have changed my mind. Not everyone may agree; such classifications are always open to interpretation. But I am now convinced that the Longgupo fossil and others like it do not represent a pre-*erectus* human, but rather one or more mystery apes indigenous to southeast Asia's Pleistocene primal forest. In contrast, *H. erectus* arrived in Asia about 1.6 million years ago, but steered clear of the forest in pursuit of grassland game. There was no pre-*erectus* species in southeast Asia after all.

The Longgupo site, discovered in 1984, lies 20 kilometres south of the Yangtze River in eastern Sichuan. At the beginning of the Pleistocene (1.8 million to 10,000 years ago) this cave sat near the northern range of a subtropical forest as rich with life as any in contemporary Africa. Unsurprisingly, the mammalian fossils dug up from Longgupo belonged to the subtropical *Stegodon–Ailuropoda* fauna found throughout the subtropical forested region south of China's Qinling Mountains (see map). The name comes from two common members — the extinct elephant-like *Stegodon* and the bear-like giant panda, *Ailuropoda*. It includes primates such as the extinct giant ape *Gigantopithecus*, as well as the ancestors of the living orangutan (*Pongo*) and gibbon (*Hylobates*).

But Longgupo also yielded a mystery jaw fragment, including the fourth premolar and the first molar. Although obviously primate, the worn enamel surfaces made precise classification difficult. Some had called it an ape whereas others saw an early human. In 1992, colleagues and I were invited to Longgupo to provide a reliable age determination and to help understand the palaeoanthropology.

The 1.8-million–2-million-year-old jaw was smaller than that of any known orangutan, living or extinct. We also compared it with primate dental fossils from the site of Lufeng, in neighbouring Yunnan province. *Lufengpithecus* was of the right size and general morphology, but the age was wrong: Lufeng and similar sites belonged to the late Miocene period, about 7 million–9 million years ago. Some possible stone tools found at the site seemed to support a human classification. Asian *H. erectus* was the obvious possibility, but the size, tooth proportions and root structure were not quite right. Dissatisfied with the usual regional comparisons, we looked to Longgupo's possible links with early African humans such as *H. habilis*, whose Great Rift Valley fossils are as old as 2.3 million years. Our *Nature* announcement[1] thus presented the Longgupo jaw as a newcomer to the *Stegodon–Ailuropoda* fauna: an African hominin more primitive than *H. erectus*.

> "More than a decade later, with some distance from the subject, the teeth looked distinctly more ape-like."

Pre-*erectus* claims

We weren't the first or last to suggest that a pre-*erectus* African hominin migrated to east Asia. In the 1940s and 1950s pre-*erectus* African claims were made for fossils from Sangiran, on Java, Indonesia. Early in Sangiran's long history of *H. erectus* discoveries, a couple of massive jaws seemed similar to those of South African australopiths — they were coined 'Meganthropus'. But as more fossils were discovered at Sangiran, it became clear that the *Meganthropus* jaws were merely a local variant of *H. erectus*.
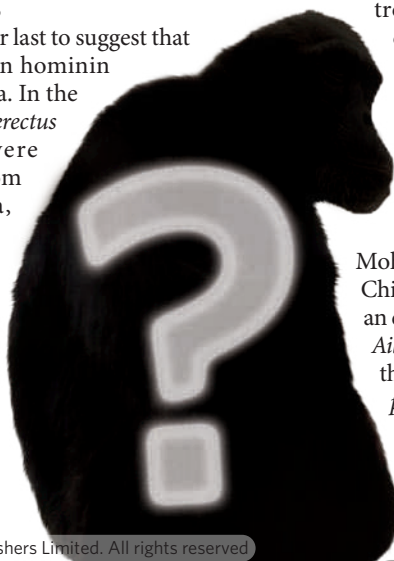
Just this year, claims for a pre-*erectus* African in Asia have also surfaced to explain the evolution of Indonesia's *Homo floresiensis*, popularly known as the Liang Bua 'hobbit'. Discovered in 2003, and dated to just 95,000 to 17,000 years ago, the Liang Bua skeleton is a diminutive species significantly different from all other known humans. The discoverers proposed that the diminutive *H. floresiensis* evolved from a southeast Asian *H. erectus* group that became isolated on Flores: faced with limited resources, the *erectus* group dwarfed to match the small-island conditions. However, recent studies of Liang Bua wrist and foot bones reveal primitive anatomies reminiscent of *H. habilis* or *Australopithecus*, again leading some to propose a pre-*erectus* African origin for the species. The problem is that no comparable wrist or foot bones are known for *H. erectus*, making it impossible at this time to exclude a local variant of *H. erectus* as the ancestor of the Liang Bua 'hobbit'.

So our claim of a pre-*erectus* African hominin living in east Asia fell into a long line of such arguments. It was met with healthy scepticism. We were first faced with the response that Longgupo was an orangutan, but we were able to show that the two teeth lay significantly outside the orangutan range of variation[2,3]. Later, we had to field a serious proposal that Longgupo belonged to *Lufengpithecus*[4,5]. Although the age disparity remained troubling, the dental similarities could not be denied. I began to imagine a mystery ape as a possible solution to the problem.

Then in spring 2005, I met with Wang Wei, director of the Guangxi Natural History Museum, to examine his collection of 33 primate teeth from Mohui cave in Bubing Basin, south China. Wang's excavations produced an excellent sample of the *Stegodon–Ailuropoda* fauna. Quickly I could see that 15 teeth were those of *Gigantopithecus*, and 10 were probably *Pongo*. The remaining eight specimens did not fit with any known east Asian Pleistocene primate.

W. FERNANDES

ON THE HOMININ TRAIL
Perhaps Pleistocene humans stayed clear of the subtropical forest, preferring grasslands instead.

Legend:
○ *Homo erectus* sites
▪ Mystery ape sites
Pleistocene coastline
Stegodon-Ailuropoda fauna region/forest (faded area shows unknown boundary)

Some 15 years earlier, I had worked hard to show that *Gigantopithecus* had crossed paths with *H. erectus*; I wrote a book in 1990 proposing this relationship (*Other Origins: The Search for the Giant Ape in Human Prehistory*) and a few years later had documented evidence of the species' co-existence. In my mind the two were firmly linked. But more than a decade after the discovery, with some distance from the subject, the teeth in Wang's lab looked distinctly more ape-like than hominin.

Teething problems

Without the assumption that *Gigantopithecus* and *H. erectus* lived together, everything changed: if early humans were not part of the *Stegodon–Ailuropoda* fauna, I had to envision a chimpanzee-sized ape in its place — either a descendant of *Lufengpithecus*, or a previously unknown ape genus. The Mohui mystery teeth surely belonged to an unknown ape, as did Longgupo, and other human-like teeth often identified from similar cave fossils. Although I no longer consider the Longgupo jaw to be human, the two stone tools still stand as described. They must have been more recent additions to the site.

The mystery ape concept is bolstered by looking at definitive *H. erectus* finds in east Asia. Our knowledge comes mainly from two sites: Zhoukoudian near Beijing, which lies well north of the primal forest, and Sangiran in Java, which lies well south of it. Each site represents hundreds of thousands of years of *H. erectus* occupation: Sangiran beginning as early as 1.6 million years ago, Zhoukoudian beginning about 780,000 years ago[6]. Neither site preserves *Stegodon–Ailuropoda* fauna or mystery ape teeth. *Homo erectus*, it seems from this perspective, hunted grazing mammals on open grasslands, and did not or could not penetrate the dense subtropical forest. In fact, there is no record of early hominins living in tropical or subtropical forested environments in Africa or Asia.

In resolving the mystery, two other Asian sites come to mind: Jianshi (Hubei province, China) and Tham Khuyen (Lang Son province, Vietnam). At both sites, teeth labelled variously as *Australopithecus*, *H. erectus* and *Meganthropus* are most likely to be the mystery ape instead. Others have come to similar conclusions[7]; a 2009 paper identifies a tooth from Sanhe Cave (Chongzuo, Guangxi province,

China) as belonging to an unidentified ape[8].

In this call to reassess historical assemblages, it is worth remembering the story of '*Hemanthropus*'. Legendary fossil collector Ralph von Koenigswald created this hominin taxon in 1957, based on isolated fossil teeth found in apothecary shops across southeast Asia. Von Koenigswald viewed *Hemanthropus* as a distant relative of African *Australopithecus*. Later research revealed that these were worn or atypical orangutan teeth and *Hemanthropus* was quickly abandoned. But, had von Koenigswald actually discovered evidence of the mystery ape? In October 2005, I examined the original *Hemanthropus* collection. Among the many worn orangutan teeth I found several small ape teeth that very closely resembled the mystery ape teeth from Mohui. Perhaps von Koenigswald was the first to lay hands on the mystery ape.

The question remains: is there only one mystery ape or possibly more? It seems that there was as much diversity of apes in the southeast Asian Pleistocene as in Africa today. In modern Africa there is one large ape (the gorilla) and two smaller apes (the chimpanzee and bonobo); in Asia during the Pleistocene and recent times, we have one very large ape (*Gigantopithecus*), one large ape (the orangutan), at least one smaller ape (mystery ape) and finally a tiny ape (the gibbon).

The next step is to consider the mystery ape fossils as a group and see how they fit into the evolutionary history of the range of southeast Asian apes. Wang will head up this team effort, along with Chinese and international colleagues, including myself. Museum collections holding potential mystery-ape evidence will be examined, including those in Hanoi, Jianshi, Beijing and Frankfurt. Wang's ongoing excavations at cave sites in Guangxi's Bubing Basin are yielding new evidence with every passing day. Possibly, there will be a chance to announce a new southeast Asian fossil ape in some future issue of this journal.   ▪

Russell L. Ciochon is chair of anthropology at the University of Iowa, Iowa City, Iowa 52242, USA. This Essay is based on a contribution to the book *Out of Africa I: Who, When and Where?* (eds, Fleagle, J. G. *et al.* Springer, 2009).
e-mail: russell-ciochon@uiowa.edu

1. Huang, W. *et al.* *Nature* **378,** 275–278 (1995).
2. Schwartz, J. H. & Tattersall, I. *Nature* **381,** 201–202 (1996).
3. Huang, W. *et al.* *Nature* **381,** 202 (1996).
4. Wu, X. *Acta Anthropol. Sin.* **19,** 1–10 (2000).
5. Etler, D. A. *et al.* *Hum. Evol.* **16,** 1–12 (2001).
6. Ciochon, R. L. & Bettis, E. A. III *Nature* **458,** 153–154 (2009).
7. Schwartz, J. H. *et al. Anthropol. Pap. Amer. Mus. Nat. Hist.* **76,** 1–24 (1995).
8. Jin, C. *et al. Chinese Sci. Bull.* **54,** 788–797 (2009).

BOOKS & ARTS

A fresh take on food

A slew of publications examines our changing attitudes to the things we eat, so what lies behind our need for mutant maize or locally grown organic food, asks **Jascha Hoffman**?

"What's in the fridge?" may not seem a weighty question. But food is one of our oldest and most advanced technologies. And, as two new books and a documentary film show, we all have a stake in what we eat.

Over the centuries, armies and empires have stood and fallen on the strength of their provisions. In *An Edible History of Humanity*, Tom Standage, business editor of *The Economist*, does an admirable job of showing the "invisible fork" behind the fate of nations. Attributing "social transformation … geopolitical competition … military conflict and economic expansion" to the cultivation of a handful of grains and meats, he explains how farming allowed bands of humans to settle and grow into societies. Early civilizations bloomed after the domestication of key plants: wheat and barley in the Near East, rice and millet in Asia, and maize and potatoes in the Americas. Later, the promise of spices lured European explorers around the globe, where they made their fortunes on transplanted crops such as sugar cane.

Standage is no cheerleader for agricultural innovation. He suggests that maize (corn), which was bred "from a simple grass into a bizarre, gigantic mutant that can no longer survive in the wild", may have exploited us as much as we did it. The invention of farming might even have been "the worst mistake in the history of the human race" as it meant more work and a more restricted diet than was available from hunting and gathering. Yet Standage is bullish on recent efforts to master our crops, praising the pioneers of nitrate fertilizer and high-yield grains that caused a sharp spike in food production. He claims not to place blind faith in biotechnology, but believes that our best chance of meeting the global food crisis is to engineer hardier soya beans.

After growing food, we must get it to the

With the advent of refrigerators, cold food became a sign of prestige.

An Edible History of Humanity
by Tom Standage
Walker & Company/Atlantic Books:
2009. 288 pp./368 pp. $26/£19.99

Fresh: A Perishable History
by Susanne Freidberg
Belknap–Harvard University Press:
2009. 416 pp. $27.95, £20.95,
€25.20

Food, Inc.
Film directed by Robert Kenner
Opened on 12 June 2009
See www.foodincmovie.com

table before it goes bad. In *Fresh*, Susanne Freidberg, a historian at Dartmouth College in New Hampshire, chronicles how expectations about beef, fish, milk, eggs, fruit and vegetables have shifted over the past century. Freshness means more than the absence of biochemical decay. It is bound up with our notions of purity, nutrition and beauty. And these ideas have adapted to the rise of a technology that most of us now take for granted — refrigeration.

For most of human history, the only way to keep unpreserved meat from spoiling was to haul in ice from a nearby mountain range. After 1876, when the first refrigerated steamship brought tonnes of fresh beef across the Atlantic, refrigeration caught on quickly in the United States. Small-town butchers were ruined by the chilled railcars of the early Chicago meat packers. By the 1920s, when housewives were clamouring to replace their iceboxes with Kelvinators and Frigidaires, cold food had become a mark of prestige. Consumers stopped expecting 'fresh' to mean just picked, caught or killed.

Coldness became a sign of freshness.

Freidberg is perceptive about how consumer attitudes responded to these mechanical advances and the marketing campaigns that surrounded them. But she gives only a small sample of the innovations that are redefining freshness today — from NatureSeal, a calcium-citrate formula that can keep a cut apple looking good for a month, to the "tamper proof, laser-coded, traceable egg" promised by a company called Eggfusion. She keeps admirably cool when discussing the global trade in baby vegetables in which some of the most perishable crops — haricots verts from Burkina Faso, mini courgettes from Guatemala, baby sweetcorn from Zambia — travel halfway around the world to market. But her culinary sense of betrayal is unmistakable when she laments that the supermarket doctrine of "permanent global summertime … has largely destroyed the seasonality of fresh produce and with it, many would argue, its taste."

Journalists Michael Pollan and Eric Schlosser, who star in a documentary by Robert Kenner titled *Food, Inc.*, broaden the attack on 'Big Food' on the basis of health, labour and the environment. After conceding the efficiency of the enormous slaughterhouses that churn out hundreds of thousands of identical pork chops each day, the film brings a litany of complaints against factory farming. It shows how US grain subsidies have led to the overconsumption of corn syrup and grain-fed meat; and how a 'revolving door' between agribusiness and the US Food and Drug Administration has led to a sharp decline in livestock inspections despite persistent bacterial scares. The film depicts how big companies can ruin small farmers by suing them for replanting their patented seeds, and tracks the growing backlash among a handful of agricultural rebels, including a rancher from North Carolina who prefers to slaughter his own grass-fed animals and a dairy tycoon who believes he can "save the world" by selling more high-end yogurt to Wal-Mart shops. If there were any doubt about the film-makers' sympathies, the closing credits counsel viewers to buy healthy, organic and local products.

One expects a nod of agreement from Freidberg, who approves of 'locavores' —

proponents of locally sourced foods, such as Pollan and restaurateur Alice Waters (see Q&A, below) — because "less transport and storage of fresh foods saves vitamins as well as energy". But Standage, poring over the numbers, believes that the current obsession with calculating 'food miles' is misguided. Transporting food may take less energy than growing or cooking it. So the carbon footprint of an English lamb chop, raised on energy-intensive maize feed, can be larger than that of a grass-fed one imported to Britain from New Zealand. If consumers want to save energy, he suggests, they should consider leaving the lid on the pan.

Arguments over food will continue as long as humans survive to eat it. As Standage contends, "every thing that every person has ever done … has literally been fueled by food". That fact of history is unlikely to change. ∎

Jascha Hoffman is a writer based in New York.
e-mail: jascha@jaschahoffman.com

Q&A: Education from the ground up

When **Alice Waters** founded Chez Panisse restaurant in 1971, she used fresh ingredients from local suppliers and sparked a culinary revolution in Berkeley, California, that has spread worldwide. For the past decade she has been taking that revolution into education. Waters talks about teaching science in the garden, and the true cost of a school lunch.

Why did you start teaching children to grow vegetables?

A local school principal called me to see if engaging children in gardening and cooking could change the way they ate. Having worked in my restaurant for 25 years, I realized I shouldn't be alone in tapping such a pristine source of local food. At a jail in San Francisco, I had seen inmates so engaged in gardening that they did not want to leave. I knew there was something deep in the idea of putting your hands in the earth. That is what we discovered in my Edible Schoolyard programme: if kids grow it and cook it, they will want to eat it.

How do you teach science through gardening?

You can weave food into the curriculum in the most imaginative ways — the growing of it, the cooking of it, the eating of it. You can get young kids to count varieties of beans instead of buttons. When the kids go out in the garden, they measure the beds and calculate the number of seeds that have to be planted. They work on irrigation systems. They see how quickly the worms decompose the vegetables. It's science and they think it is fun. I was a teacher before I started my restaurant.

What's wrong with most school lunches?

In the United States, the schools get the least healthy foods, and there is not enough money to pay real cooks. You have a deadly combination of surplus commodities that are basically fast food and reimbursement that's so small that you cannot prepare fresh salads and fruits. The cost? An obesity epidemic, the pollution of the environment and the destruction of our culture.

Should governments tell us how to eat?

We believe that eating is our own private turf. But if other people are eating poorly,

Alice Waters takes learning outdoors with her hybrid programme of gardening and science.

T. HEINSER

it raises the rates of everyone's health care. This is a global issue that governments must regulate. We have to decide what constitutes food. In France they have national health care, so they measure and weigh children to tell them how many calories they should have for lunch. We need 'edible education' to start at an early age. The government should pay up front for that, instead of cleaning up later with health care.

Have you discussed this with the administration of Barack Obama?

It has been my goal since 1992 to encourage every US president to plant a victory garden in the spirit of President Thomas Jefferson. I met Michelle Obama last summer, and I know people in Congress are talking about linking farms and schools. Hillary Clinton is a big proponent too. I call this a stimulus package. You nourish all children with food that is good, clean and fair. That means buying it from local sources. It is a moral issue. It cannot wait another day.

What other food-education work is going on outside of California?

We're concentrating now on New Orleans. Because it is so fertile and so in need after Hurricane Katrina we wanted to help, and in three years the programme has blossomed. It is expensive — it costs hundreds of thousands of dollars to hire the best teachers and to take care of visitors. The Yale University Sustainable Food Project of the past six years is also a success. It has integrated an organic farm that grows 250 varieties of fruits and vegetables into the curriculum. Hundreds of students volunteer to work in the gardens and they are demanding fair-trade coffee and grass-fed beef all over campus. I'm going to Harvard University soon to see whether we can change that old food system too.

What about countries that are struggling to feed their children?

I've been in touch with a teacher in Ghana who said, "For our ultimate survival, we have to engage the children in the cooking of their own food." Former United Nations secretary-general Kofi Annan has also talked about the green revolution in Africa. People in Africa haven't come to this from taste and beauty, as I did, but from the ideas of hunger and necessity. But we all arrive at the same place. We need to feed everybody on this planet. ∎

Interview by Jascha Hoffman, a writer based in New York.
e-mail: jascha@jaschahoffman.com

Sky shoots: Soonil Kim's proposed aerial vineyard for London.

S. KIM, ARCHITECTURAL ASSOCIATION

Seeds of an edible city architecture

Global warming and food shortages are renewing interest in urban agriculture, finds **John Whitfield**.

In Brad Bass's imagination, a visit to the salad bar could become less a chance to load your plate and more a foraging expedition. "I have this vision of restaurants putting up walls, and people picking their own vegetables right off the wall instead of pulling them out of the tub," says Bass, a geographer at the University of Toronto in Canada.

Something akin to Bass's vision has just been displayed in London. *Urban Agriculture Curtain*, an indoor, vertical vegetable patch, was installed by architectural firm Bohn and Viljoen at an exhibition on urban agriculture that ran at the city's Building Centre last month. Visitors to *London Yields* were not allowed to pick their own, but they could eat the exhibit's produce in the centre's cafeteria. Some of the most imaginative ideas displayed came from students, such as Soonil Kim's proposal for an aerial vineyard — with vines growing in barrels suspended from cables — in London's King's Cross neighbourhood. "Students look at the new ground, and they've pre-empted a lot of the current interest," says curator Jackson Hunt.

Several exhibitions worldwide are surfing a wave of activity in integrating plants and architecture. People have been clothing buildings in vegetation for millennia, at least since the Hanging Gardens of Babylon were constructed around 600 BC. The Vikings insulated their roofs with turf, and the shopkeepers of Pompeii in Italy shaded their balconies with vines.

London Yields: Urban Agriculture
Main Gallery, The Building Centre, London
9 April until 30 May 2009

Vertical Gardens
Exit Art, New York City
28 March until 6 June 2009

Radical Nature: Art and Architecture for a Changing Planet 1969–2009
Barbican Art Gallery, London
19 June until 18 October 2009

But growing populations, expanding cities, concerns about food security and climate change have given the idea a new prominence. *Vertical Gardens*, an exhibition held this spring at New York's Exit Art gallery, highlighted projects for growing plants on and in modern buildings. Examples already built include Patrick Blanc's green wall on the Quai Branly Museum in Paris. Bohn and Viljoen presented a conceptual design for a high-rise tower clad in 'vertical fields', and displayed a mock-up of a balcony offering a view through the branches of an espaliered fruit tree. Dickson Despommier of Columbia University, New York, presented his idea for a vertical farm — a 30-storey tower housing enough animals and plants to produce food for 50,000 people.

Further along the spectrum from architecture to art is a forthcoming show at the Barbican Art Gallery in London. *Radical Nature*, which opens on 19 June, comments on humans' relationship with the environment rather than suggesting explicit ways to improve it. Attempts to re-imagine cities are not new: Agnes Denes, in her 1982 work *Wheatfield — A Confrontation*, photographs of which are displayed at the Barbican, planted almost a hectare of New York's Battery Park with wheat in an effort, she wrote, "to call attention to our misplaced priorities and deteriorating human values". Events on guerrilla gardening and farmers' markets accompany the exhibition.

Green roofs have been installed on new buildings for decades, and the technology and business surrounding them are mature, says Bass, who has studied their environmental effects. Plant-covered roofs, such as that of the refashioned California Academy of Sciences in San Francisco (see *Nature* **455**, 466; 2008), can cut energy bills, reducing the need for both heating and cooling. They mop up pollutants and cool the air above the building, reducing the heat islands that develop in urban areas. And they slow the rate of water run-off, which eases the burden on city drains, thus reducing the risk of floods and pollution of waterways. Well-chosen roof plants may enhance urban biodiversity. If the roof is accessible and can take the weight of the deeper soil needed, you can even grow food up there.

E. DE LA OSA/REUTERS

Grass roots: Havana's urban gardens yield tonnes of fresh vegetables for Cubans every year.

Vertical gardens likewise provide insulation, soundproofing and physical protection to the building's fabric, and they can shade windows, further reducing the need for energy-hungry climate control. They are, however, more complex to build and design than green roofs. Growing plants on a wall, as opposed to trailing them over it, requires a hydroponic system. The plants are rooted in plastic or plant fibres rather than soil, and water and nutrients are pumped into the medium. Although his *Urban Agriculture Curtain* uses this technology, architect André Viljoen questions the environmental benefits of hydroponic and indoor-grown food. "I'm sceptical about the amount of chemicals and energy used in hydroponic systems," he says. "And once you start heating or lighting these things, the environmental benefit goes out of the window."

A better option, says Viljoen, is to weave organic farming into the fabric of a city, inserting market gardens into their patchwork of industrial, residential, recreational and empty land. He has studied Cuba's use of such methods to feed its people after the collapse of Soviet aid. The government encouraged private and communal vegetable growing in cities, and developed a growing system dubbed *organopónico*, which replaced petrochemical fertilizers with organic ones such as sewage sludge. In 1997, Havana's system produced almost 21,000 tonnes of vegetables; in 2005, that had risen to 272,000 tonnes, and the project has become a flagship for similar efforts worldwide. Viljoen believes that, in the temperate developed world, 30% of a city's fruit and vegetables could be grown within its borders. By reducing food miles, this could yield big cuts in carbon emissions. But for larger reductions, he adds, you would also need changes in diet, as most of the emissions due to food come from the meat industry.

What is needed now, say both Hunt and Viljoen, are pilot projects to test the large-scale potential of green buildings and urban agriculture, which also take into account the health, amenity and aesthetic benefits provided. "Any one argument looked at on its own tends not to be strong enough," says Viljoen. "But collectively you can make a very strong case." Such schemes are more likely to come from the retrofitting of existing spaces and buildings than from high-tech, high-concept projects, says Hunt, as people green their own environments and pressurize local government to do the same. "There's a groundswell of interest in the subject, but developers are very conservative by nature," he says. "If it's going to happen it's going to be community-driven." ■

John Whitfield is a writer based in London and author of *In The Beat of a Heart*.
e-mail: ja_whitfield@hotmail.com

Sustainable fashion

The sign beside the thick, soft, creamy wool rug says, "Please do not touch." Naturally, I want to roll on the rug and wrap myself in it. Each of its 11 patterned hexagonal panels was knitted from the wool of a Panama sheep that had grazed on pasture untainted by pesticides at Lava Lake Ranch, Idaho. Made by Dutch designer Christien Meindertsma using extra-large needles, the certified-organic rug forms part of *Design for a Living World*, an exhibition now on show at the Smithsonian's Cooper-Hewitt, National Design Museum in New York.

Organized with The Nature Conservancy, a global conservation group based in Arlington, Virginia, the exhibition aims to raise awareness of "the impact and promise of sustainable sourcing". Ten prominent designers were invited to create an object — a chair, a dress, a necklace — using sustainably grown and harvested materials from some of the world's most beautiful and fragile places. The sale of such objects could help to provide a livelihood for local communities in these areas — many of which face threats from over-development and deforestation — and also emphasizes how designers influence what we buy.

Fashion designer Isaac Mizrahi created a dress covered in creamy-white, sequin-like leather paillettes made from the skin of wild-caught salmon from southwest Alaska. Jewellery designer Ted Muehling fashioned bracelets and delicate flowers out of 'vegetable ivory' extracted from the seeds of the ivory nut palm tree on the Micronesian island of Pohnpei in the western Pacific Ocean. And Maya Lin — famous for designing the powerful Vietnam Veterans Memorial in Washington DC — crafted a simple bench from a single red maple. The tree was harvested from woods certified by the Forest Stewardship Council around the Upper St John River in Maine.

Some items are both elegant and useful. Israeli designer Ezri Tarazi constructed "a bamboo forest inside your living room" — towering bamboo stems fitted with clothes hooks, compact-disc racks, wine-bottle holders and lights that glow through ping-pong balls slotted in holes in the stalk. The bamboo, harvested from China's Yunnan Province, can grow a metre a day and requires little water to flourish in its natural habitat; although, a caption warns, exploding demand threatens some established forests that are being cleared to make way for bamboo plantations.

Second skin: Isaac Mizrahi used discs of salmon leather to adorn a dress.

M. STROH

Less functional are the odd objects shaped by Dutch designer Hella Jongerius out of chicle latex, extracted from the chicozapote tree in the Mayan rainforest on Mexico's Yucatán Peninsula. Once the basis for chewing gum, the elastic, viscous goo refuses to lend itself to any obvious purpose. Jongerius refers to it affectionately as "an alien in the house", and sticks strings or blobs of it around the necks of vases.

Jongerius's whimsical project raises questions as to the practical significance of the exhibits and whether they can translate into long-term, widespread and commercially viable uses for sustainable materials and practices. It is hard to imagine the average shopper investing in Lin's graceful bench. But Mizrahi believes that big opportunities await fashion designers who recognize that protecting the environment makes economic sense.

Salmon skin, for example, is normally discarded as a waste product, yet it is a valuable and resilient material that can be turned into shoes, belts and bikinis. It requires less-toxic chemicals for tanning than mammal hides because fish scales are easier to remove than hair. "People think of salmon skin as something you peel off your food; in fact it's this beautiful substance," Mizrahi says in a video accompanying his exhibit. He adds, "The fashion business is crazily competitive. All of a sudden it's going to occur to these greedy people that they can make a lot of money if they conserve." ■

Josie Glausiusz is a writer based in New York.
e-mail: jg@planetjosie.net

Design for a Living World
Cooper-Hewitt, National Design Museum, New York City
Until 4 January 2010

NEWS & VIEWS

ORGANOMETALLIC CHEMISTRY

Charting the course of catalysis

Gregory L. Hamilton and F. Dean Toste

With palladium catalysts, chemists can manipulate ordinarily inert carbon–hydrogen bonds to build useful molecules from simple building blocks. How the catalysts guide this process has just become a bit clearer.

Carbon–hydrogen bonds are one of the most common types of linkage in organic molecules. But they do not react under most chemical conditions, and so are not generally considered to provide a useful handle for building complex molecules. Consequently, it has long been a goal in synthetic chemistry to selectively transform C–H bonds into other functional groups of atoms that can be easily manipulated. Certain transition-metal complexes can induce typically inert C–H bonds to react in a variety of ways[1], and an aim has been to find metal compounds that act as efficient catalysts so that only small amounts of them are required (the metals used are typically quite expensive). A further challenge is that successful catalysts need to be able to pick one C–H bond out of many in a molecule.

A significant advance came with the discovery[2] of a series of C–H bond transformations that require only a catalytic amount of palladium complex, together with an oxidizing agent, and that proceed with high yield and selectivity. By varying the type of oxidant used in conjunction with the palladium catalyst, it is possible to form new carbon–oxygen, carbon–nitrogen, carbon–halogen and carbon–carbon bonds (Fig. 1a). The precise mechanism by which palladium performs this remarkable feat has been the subject of intense study and controversy[3–6].

Writing in *Nature Chemistry*, Powers and Ritter[7] throw their hats into the ring with a bold proposal about that mechanism. Analysis of how molecules get from point A to point B is, of course, complicated by the fact that we cannot, in general, directly observe molecules interacting. Instead, chemists must look at intermediates along the pathway from starting material to product to piece together a kind of molecular road map. Reaction intermediates are often unstable, making complete analysis difficult. Nonetheless, Powers and Ritter have succeeded in isolating and determining the structures of key intermediates of a particular palladium-catalysed process for replacing C–H bonds. In doing so, they shed light on this broad class of transformations in a way that should inspire future studies and maybe facilitate the development of more powerful catalysts.
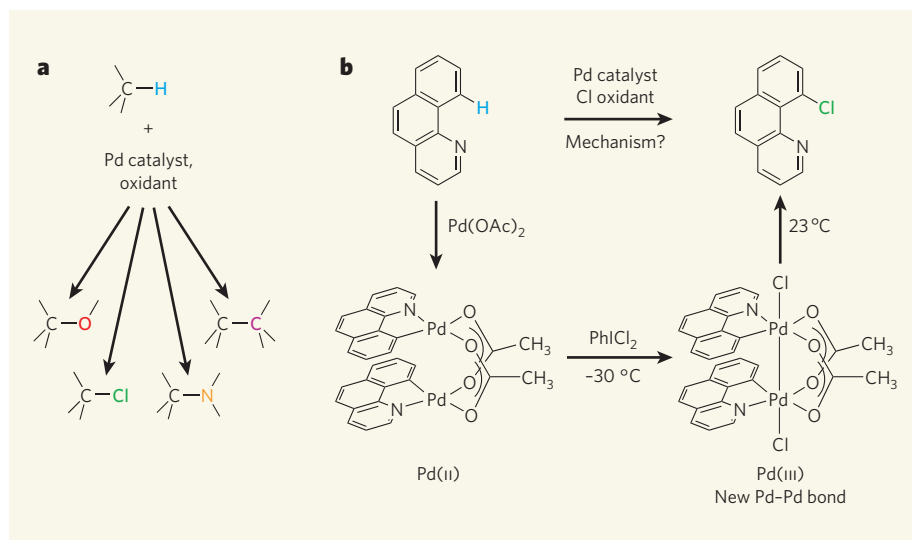
Figure 1 | Palladium-catalysed transformation of carbon–hydrogen bonds. a, By using a palladium catalyst in conjunction with a suitable oxidant, normally unreactive C–H bonds can be replaced with a variety of useful groups. This type of process can greatly simplify construction of pharmaceutical molecules and biologically active natural products. **b,** Powers and Ritter[7] performed a step-by-step analysis of a palladium-catalysed reaction that converts C–H bonds to C–Cl bonds. Insertion of palladium acetate, Pd(OAc)$_2$, into a C–H bond produces an intermediate with two palladium atoms each in the (+2) oxidation state, as expected. Surprisingly, low-temperature treatment with the chlorine-based oxidant iodobenzene dichloride, PhICl$_2$, results in a complex in which the two palladium atoms are bonded to one another and each has been oxidized by one electron. These Pd(III) intermediates have not previously been shown to take part in catalytic reactions. Finally, the isolated complex releases chlorinated product after being warmed to ambient temperature.

When considering reaction mechanisms, chemists are greatly concerned with the flow of electrons from one atom to another. To keep track, they assign an oxidation state to each atom as a bookkeeping measure loosely associated with the atom's charge. Oxidation states are particularly useful for predicting and rationalizing the reactivity of metal complexes. In reactions where palladium is a catalyst, the metal usually exists in the zero or (+2) oxidation states — denoted Pd(0) or Pd(II) — and often cycles between the two.

In a proposal initially received with scepticism, Sanford and colleagues suggested that Pd(IV) species may play a central part in oxidative C–H bond transformations. They have accumulated substantial evidence[3–6] for this theory by independently synthesizing Pd(IV) complexes related to the proposed reaction

intermediates and showing that the prepared palladium species proceed to the desired product. Detailed investigation has provided a high level of understanding of product formation from Pd(IV). A remaining question, however, is whether the synthesized complexes are close enough approximations of actual intermediates in the catalytic reactions, because different conditions are used in each case. In other words, if the route to a destination is changed, you may well expect to stop at a different place along the way, even though the endpoint is the same.

Powers and Ritter[7] address this issue by employing conditions very close to those used in a palladium-catalysed C–H bond halogenation reaction (Fig. 1b). By controlling the reaction at a low temperature, the authors were able to isolate an intermediate palladium complex and determine its structure by X-ray

crystallography. Rather than finding a Pd(IV) species, they observed a complex containing two palladium atoms bonded to each other, each in the (+3) oxidation state. This type of Pd(III)–Pd(III) bonded species (described as 'bimetallic') is quite rare and has not been previously shown to feature in catalytic reactions. By analysing the rate at which the isolated complex formed product, the authors showed that the final reaction step probably proceeds directly from the observed Pd(III) intermediate, thereby highlighting its importance in the catalysis.

Taken together, Powers and Ritter's results strongly suggest that bimetallic Pd(III) species

are intermediates in the palladium-catalysed oxidative transformations of C–H bonds presented in their study. Their report certainly removes part of the veil concealing the complicated molecular picture of this class of reaction. Naturally, the issue of the exact mechanism for each reaction is far from settled, and the relevance of previous proposals involving Pd(IV) intermediates should not be discounted. Future studies could decipher which reaction parameters dictate the preferred pathway for different processes; furthermore, the newly observed reactivity of the Pd(III) complexes should spark investigations into more reactions involving these remarkable species. ■

Gregory L. Hamilton and F. Dean Toste are in the Department of Chemistry, University of California, Berkeley, California 94720-1460, USA. e-mail: fdtoste@berkeley.edu

1. Labinger, J. A. & Bercaw, J. E. Nature 417, 507–514 (2002).
2. Dick, A. R. & Sanford, M. S. Tetrahedron 62, 2439–2463 (2006).
3. Dick, A. R., Kampf, J. W. & Sanford, M. S. Organometallics 24, 482–485 (2005).
4. Dick, A. R., Kampf, J. W. & Sanford, M. S. J. Am. Chem. Soc. 127, 12790–12791 (2005).
5. Hull, K. L., Lanni, E. L. & Sanford, M. S. J. Am. Chem. Soc. 128, 14047–14049 (2006).
6. Whitfield, S. R. & Sanford, M. S. J. Am. Chem. Soc. 129, 15142–15143 (2007).
7. Powers, D. C. & Ritter, T. Nature Chem. advance online publication doi:10.1038/nchem246 (2009).

MALARIA

The gatekeeper revealed

Sarah B. Reiff and Boris Striepen

A molecular machine used by the malaria parasite to export its protein armoury into the host cell has at last been identified, providing researchers with a potentially invaluable therapeutic target.

Malaria is a devastating infectious disease that claims around one million lives each year. The most severe form is caused by a unicellular parasite called *Plasmodium falciparum* — a formidable foe that has repeatedly frustrated attempts to eradicate it. *P. falciparum* has many survival tricks: it constantly varies the surface proteins presented to the host[1], thereby confounding the host immune system, and it rapidly develops resistance to drugs[2]. In this issue (page 945), de Koning-Ward and colleagues[3] get to the bottom of yet another of the parasite's unsettling talents, the ability to inject hundreds of its own proteins into the host cell.

The invasion of the host cell by an army of parasite proteins is probably one reason why *P. falciparum* can survive and develop in a cell that few other pathogens have managed to conquer, the red blood cell (RBC). Streamlined for a single purpose — the transport of oxygen — RBCs are cellular deserts, lacking many of the amenities usually enjoyed by intracellular pathogens, such as a nucleus and a vesicular transport system. *P. falciparum* masters this challenge by extensive remodelling of the RBC, and the parasite's injected proteins are probably the agents of these changes. One such change is the formation of knobs on the RBC surface that behave like molecular Velcro, resulting in adhesion of infected cells to the lining of capillary blood vessels[4]. By hiding out in the periphery of the circulatory system in this way, infected cells seem to avoid elimination in the spleen. The consequences for the host are dire — the massive accumulation of infected RBCs in the

capillary beds of the brain and kidney can lead to organ failure and ultimately death.

The mechanism used by *P. falciparum* to export proteins to the RBC cytoplasm has been hotly pursued by malaria researchers. Understanding this process could aid the development of interventions that block the

Figure 1 | Exporting malaria proteins across the vacuole. *Plasmodium falciparum* parasitizes red blood cells (RBCs) and resides in a specialized vacuole formed during host-cell invasion. A new study[3] suggests that the parasite installs a protein complex (PTEX) in the vacuole membrane that enables it to export hundreds of parasite proteins into the RBC cytoplasm. Maurer's clefts and knobs are structures found only in infected RBCs, and their formation and function probably depend on the activity of the exported parasite proteins.

parasite's growth or limit the severity of the disease. In the RBC, *P. falciparum* resides inside a vacuole, and exported proteins must overcome several physical barriers en route to their destinations. These barriers include the parasite membrane, the vacuole membrane and, for some proteins, the RBC membrane. Proteins exported from the parasite into the vacuolar space follow the typical secretion pathway that exists in all eukaryotic cells (cells with a nucleus and membrane-bound organelles). But how these proteins cross the vacuole membrane has remained a mystery. The identification[5,6] of a specific 'export address tag' — a sequence motif at the amino terminus of *Plasmodium* proteins that results in their export out of the vacuole — was a major advance, and led the authors to their current hypothesis[3]. They propose that there is a 'gatekeeper' in the vacuole membrane that reads the export tag and delivers proteins from the vacuole into the RBC cytoplasm (Fig. 1). They further assume that the gatekeeper is made up of a complex of parasite proteins that they call the *Plasmodium* translocon of exported proteins (PTEX).

To find evidence of a vacuolar membrane protein transporter, de Koning-Ward and colleagues[3] assembled a list of required characteristics of potential PTEX components: they should be present in the blood-stage parasite, be essential for parasite growth in RBCs, be secreted from the parasite, and associate with the vacuole membrane. Using these criteria and a battery of bioinformatic, genetic and biochemical approaches, they narrowed the field from hundreds of candidates to two parasite proteins, called HSP101 and PTEX150. HSP101 is an enzyme belonging to the AAA+ superfamily of ATPases, and would be able to power protein export using ATP as an energy source. PTEX150 is a newly identified parasite protein that interacts with HSP101, a feature consistent with its being a potential component of the PTEX. Biochemical experiments identified an additional three likely PTEX components. Among these, EXP2 is particularly interesting, as it is an integral membrane protein and could therefore anchor the PTEX
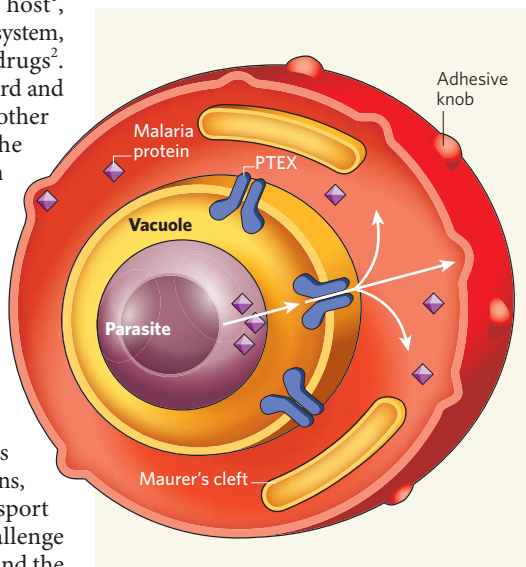
complex to the vacuole membrane and potentially function as a pore.

So de Koning-Ward and colleagues[3] seem to have identified the necessary components for the PTEX that would allow it to function as a gatekeeper. But does the PTEX actually export proteins? The authors show convincingly that it interacts with cargo proteins and, importantly, does so only if the protein carries the export address tag, which is consistent with its proposed function.

The PTEX hypothesis is not the only model proposed to explain malaria protein export. Another suggested model[7] is based on the budding of transport vesicles from the vacuole. These vesicles are then predicted to transform into membranous structures in the infected cell known as Maurer's clefts. It's possible that different sets of proteins, especially those with different destinies or physical properties, require different machinery to reach their target locations.

A third potential model arises from studies on oomycetes. These plant pathogens look like fungi, but are actually evolutionarily related to malaria parasites and also export proteins into the cells of their hosts. Intriguingly, these proteins share the *Plasmodium* export address tag, and swapping tags between oomycetes and *Plasmodium* results in export in both cases[8,9]. Oomycete researchers have made two additional observations: they noted that the address tag is also found on host proteins, and that proteins carrying the tag can gain entry to host cells in the absence of the pathogen[10,11]. This might suggest that the pathogens are subverting transport elements naturally present in host membranes.

Studying mutated versions of any of the central players thought to be involved in these models would provide crucial insights. For instance, deletion of the genes encoding the components of the PTEX complex should ablate protein export and provide conclusive proof of the PTEX mechanism. Some of the technical obstacles to generating such mutants in the malaria parasite have recently been removed[12]. The current paper[3] provides a strong mechanistic model of PTEX function and, crucially, identifies a list of excellent candidate genes to be validated in the future. ■

Sarah B. Reiff and Boris Striepen are at the Center for Tropical and Emerging Global Diseases, and the Department of Cellular Biology, University of Georgia, Athens, Georgia 30602, USA.
e-mail: striepen@cb.uga.edu

1. Scherf, A., Lopez-Rubio, J. J. & Riviere, L. *Annu. Rev. Microbiol.* **62,** 445–470 (2008).
2. Baird, J. K. *N. Engl. J. Med.* **352,** 1565–1577 (2005).
3. de Koning-Ward, T. F. *et al. Nature* **459,** 945–949 (2009).
4. Crabb, B. S. *et al. Cell* **89,** 287–296 (1997).
5. Marti, M. *et al. Science* **306,** 1930–1933 (2004).
6. Hiller, N. L. *et al. Science* **306,** 1934–1937 (2004).
7. Bhattacharjee, S. *et al. Blood* **111,** 2418–2426 (2008).
8. Bhattacharjee, S. *et al. PLoS Pathog.* **2,** e50 (2006).
9. Grouffaud, S. *et al. Microbiology* **154,** 3743–3751 (2008).
10. Birch, P. R. *et al. Curr. Opin. Plant Biol.* **11,** 373–379 (2008).
11. Dou, D. *et al. Plant Cell* **20,** 1930–1947 (2008).
12. Combe, A. *et al. Cell Host Microbe* **5,** 386–396 (2009).

BIOMECHANICS

Serpentine steps

Andrew J. Clark and Adam P. Summers

Combine theoretical modelling, friction measurements and observations of serpentine slithering. Together, they show that snakes are in effect just taking a walk even when moving at high speed.

The pendular movements of walking are dominated by gravitational forces: the body is pushed up over the planted foot, then falls back down as one foot leaves the ground and the other takes over the load. But as a walk turns to a run, inertia becomes important. Speedy locomotion in limbed animals, from mice to elephants, features this transition to a run, gallop or trot, where mass and velocity drive the animal forward.

In the unlikely event that you are chased by a snake and have to run for it, it is natural to suppose that the fast-moving snake is similarly exploiting inertial forces. In an article in *Proceedings of the National Academy of Sciences*, however, David Hu and collaborators[1] present a mathematical model that accurately predicts the biomechanics of snake locomotion — they demonstrate that even high-speed slithering is more similar to walking than running, and that it depends heavily on the frictional properties of the animal's scales.

Snakes slither on their bellies by bending from side-to-side, propelling themselves forward by pushing with their scales off microscopic bumps, or asperities, in the ground[2,3]. This force, between the scales and the ground, is of course friction; it depends only on the force pushing two surfaces together and their mutual roughness, and is characterized by a coefficient of friction. Crucially, this coefficient depends both on the surfaces involved and on the direction in which that force is applied. From cat tongues to cockleburs, we are all familiar with surfaces that show different coefficients of friction depending on the direction in which they move. Hu and his colleagues[1] show that this phenomenon is central to a snake's ability to slither swiftly.

They modelled a snake as a flexible tube with a uniform cross-section that writhed in a sinusoidal manner. If the frictional coefficients of the belly scales were made equal in all directions, the model snake thrashed about like an eel on a glass table. Clearly the model was flawed, so they moved on to measurements of friction. It is relatively straightforward to measure belly friction in snakes: all that is needed is a board and an anaesthetized snake.

In their experiments, Hu *et al.* placed the snake in various orientations on the board,

each time then tilting the board until the snake started sliding. The coefficient of friction is the tangent of the angle made between the ground and the table at which the snake first starts to move. They quickly discovered that, on very smooth boards, snake belly scales have the same frictional coefficient in all directions; but on a board covered with rougher cloth the snake slid more freely forwards than backwards. Most surprisingly, the highest friction occurred when the snake was sliding sideways.

Throwing these measured frictional values into the wriggling-sausage model yielded a faux snake that moved much like the Pueblan milk snake (*Lampropeltis triangulum campbelli*, pictured) that the authors used as a reality check. A critical parameter used in the model, called the Froude number, is a dimensionless number that appears in many contexts in physics, engineering and biology, and that in this instance can be thought of as a ratio of inertial to frictional forces. In normal terrestrial locomotion, low Froude numbers are associated with walking and higher numbers with running. In the case of snake slithering, the gravitational or frictional forces are an order of magnitude higher than the inertial forces. So, despite the perception, and even the reality, of high over-the-ground speeds, snakes are not running but walking.

The authors also noticed that as snakes slithered faster, they lifted the curved parts of their bodies away from the ground as they pushed off the ground with the uncurved parts, all of which is accomplished by redistributing body weight. Modifying the mathematical model to take the body lifting into account, they found an even better predictor of speed. The lifted-snake model was also used for calculating and visualizing the distribution, direction and relative magnitudes of frictional forces along the slithering body. The snake is maximizing its weight in the areas of its body that are subject to a sidewards force. This has the twofold effect of increasing the stability of the pushpoint and decreasing the friction elsewhere on the body.

Studies of snake biomechanics have facilitated the engineering of snake-inspired robots to perform search-and-rescue missions in

M. HARVEY/NHPA

unstable and dangerous environments[4,5]. Until recently, robots designed for such places all had wheels, which neatly offer huge differences in the friction in the fore-and-aft and side-to-side directions. Hu and colleagues' investigations[1] provide guidance on the types of material that might be used to manufacture a robot belly

that offers similar frictional properties to those of snakes. As a form of locomotion, the slither has a lot going for it. ■

Andrew J. Clark and Adam P. Summers are in the School of Biological Sciences, University of California, Irvine, California 92697, USA.
e-mails: aclark@uci.edu; asummers@uci.edu

1. Hu, D. L., Nirody, J., Scott, T. & Shelley, M. J. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0812533106 (2009).
2. Mosauer, W. *Science* **76**, 583–585 (1932).
3. Gray, J. *J. Exp. Biol.* **23**, 101–120 (1946).
4. Guo, Z. V. & Mahadevan, L. *Proc. Natl Acad. Sci. USA* **105**, 3179–3184 (2008).
5. Ostrowski, J. & Burdick, J. in *IEEE Int. Conf. Robot. Autom.* 1294–1299 (IEEE, 1996).

PLANETARY SCIENCE

Io's escape

Gerald Schubert

According to the latest study, our witnessing of the volcanic splendour of Jupiter's moon Io might just be a lucky circumstance. The odds are that the satellite will become quiescent on its escape from orbital custody.

Jupiter's moon Io is about the same size as our Moon, but the similarities end there. Its motion around Jupiter is tightly constrained by its gravitational interactions with the giant planet and with two of Jupiter's other moons, Europa and Ganymede. Io's orbital imprisonment is the cause of its spectacular volcanism[1].

But Lainey and colleagues[2], on page 957 of this issue, provide evidence that Io is loosening the bonds that hold it in its unexpectedly elliptical path around Jupiter. If it eventually breaks free, the most volcanically active object in our Solar System will become dormant.

Io's intense volcanism is driven by tidal heat

Figure 1 | Jupiter–Io tidal interaction. a, The tide raised on Io by its parent planet Jupiter — indicated by the exaggerated elliptical distortion of the body — lags behind the Io–Jupiter line at perihelion (the moon's orbital point closest to Jupiter) because Io's orbital rotation is faster than its spin at this position in its orbit. The gravitational torque exerted by Jupiter on Io's tidal bulge tends to increase Io's spin at the expense of its orbital energy, and forces Io to move inward. **b**, The tide raised on Jupiter by Io leads the Io–Jupiter line because Jupiter rotates faster than Io orbits the planet. The Jovian planet's tidal bulge pulls Io forward in its orbit and, as a result, the moon gains orbital energy and moves outward from Jupiter. Lainey and colleagues[2] show that the effect of the tide raised on Io by Jupiter wins out, so Io is moving in towards Jupiter. (Both panels depict the tidal interaction looking from above the orbital plane of Io.)

created as Jupiter's differential gravitational pull causes the satellite to deform continuously and repeatedly as it orbits the planet: Io's surface probably moves up and down by more than 10 metres over an orbit. Jupiter's varying gravitational pull and the resulting oscillations in the tides on Io occur because the satellite is in an elliptical orbit. If Io were an isolated satellite, tidal heat would circularize its orbit. But Io's orbit is elliptical, a circumstance dictated by the moon's gravitational interaction with Europa and Ganymede (an interaction referred to as a Laplace resonance). Although Io's spin and orbital periods still have equal values, Jupiter oscillates back and forth in Io's sky as Io is alternately closer and more distant from the planet.

It requires energy to tidally deform a body, and part of this energy is eventually dissipated as heat by frictional processes within the body. The energy for Io's tidal heating ultimately comes from the energy of Jupiter's rotation. Figure 1 sketches the gravitational tidal interaction of the planet and its moon. The interaction is two-way, with the planet raising tides on the moon (Fig. 1a) and the moon raising tides on the planet (Fig. 1b). These tides influence the orbital motion of the moon in opposite ways.

The tide raised on Jupiter by Io is carried ahead of the line that connects the two objects — the Io–Jupiter line — because Jupiter spins faster than Io orbits Jupiter. The gravitational attraction of the part of Jupiter's tidal bulge that is closest to Io pulls Io forward in its orbit, whereas the attraction of the part of the tidal bulge farthest from Io retards the moon's orbital motion (Fig. 1b). The effect of the part of the bulge closest to Io dominates simply because it is nearer to the moon. As a result, Io gains orbital energy and moves away from Jupiter, whereas Jupiter loses rotational energy. This is the same interaction that is responsible for the Moon receding from Earth as Earth's rotation rate decreases over time.

However, there is another effect taking place that causes Io to spiral inward towards Jupiter. The planet tidally deforms Io as shown in Figure 1a at the instant that Io passes perihelion (its orbital point closest to Jupiter). The tidal deformation lags behind the Io–Jupiter line because, at perihelion, Io's orbital rotation is faster than its spin. The gravitational pull on the part of Io's tidal bulge nearest to Jupiter acts to increase the satellite's spin to match its orbital rotation. The opposite effect occurs at aphelion (Io's orbital point farthest from Jupiter), but

the interaction at perihelion dominates. As a result, Io's spin tends to increase at the expense of its orbital motion; Io loses orbital energy, its orbital period decreases, and it moves inward towards Jupiter.

So, is Io moving towards or away from Jupiter? Which of the above effects wins out? Lainey *et al.*[2] carry out a numerical integration of Io's orbital dynamic motion and constrain their computation by astronomical observations of the innermost Galilean moons (Io, Europa and Ganymede) acquired over the 116-year period between 1891 and 2007. They conclude that Io is moving in towards Jupiter and that Europa and Ganymede are moving away from the planet. Others[3,4] have attempted the same calculation in the past, but with poorly constrained — and often contradictory — results, probably owing to approximations made in their orbital dynamical models.

The consequences of Lainey and colleagues' results[2] are manifold. First, the present global heat loss from Io can be accounted for by tidal dissipation in the moon — that is, Io is in an approximate thermal steady state, in which its radiative heat loss is being balanced by the tidally driven, frictional production of heat. Io's thermal state has until now been uncertain, with suggestions that its thermal behaviour could involve oscillatory periods of activity[5].

Second, Io, Europa and Ganymede are moving out of resonance. (The Laplace resonance is a special orbital configuration in which Europa's orbital period is twice that of Io, and Ganymede's orbital period is twice that of Europa.) Lainey *et al.* predict the rate at which the satellites' orbital periods are moving away from the 2:1 ratio (Europa:Io and Ganymede:Europa), but they do not estimate at what point the Laplace resonance will be effectively broken. If this occurs on a short timescale, say 10^8 years or less, then we have been lucky to see Io in its volcanic glory, because dormancy will be the fate of Io when the resonance is broken (S. Peale, personal communication).

Finally, the estimated tidal dissipation in Jupiter is close to the upper-bound value associated with the expansion of the orbits of Io, Europa and Ganymede from positions in the past that were closer to Jupiter to their present orbital locations. The physical mechanisms of tidal dissipation in giant planets are not well understood, and the authors' estimate, if correct, would be an important constraint on these mechanisms. ■

Gerald Schubert is in the Department of Earth and Space Sciences and the Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567, USA.

1. Peale, S. J., Cassen, P. & Reynolds, R. T. *Science* **203,** 892–894 (1979).
2. Lainey, V., Arlot, J.-E., Karatekin, O. & Van Hoolst, T. *Nature* **459,** 957–959 (2009).
3. Lieske, J. H. *Astron. Astrophys.* **176,** 146–158 (1987).
4. Aksnes, K. *et al. Astron. J.* **89,** 280–288 (1984).
5. Ojakangas, G. & Stevenson, D. *Icarus* **66,** 341–358 (1986).

INFLAMMATION

Wound healing in zebrafish

Paul Martin and Yi Feng

What is the first signal that directs the rapid influx of immune cells to a wound to stave off potential infection? A study in the zebrafish reveals an unusual but well-qualified candidate.

The rapid recruitment of white blood cells to a wound site after tissue damage is essential to prevent the entry of microorganisms into the breach and to help to coordinate wound closure. These immune cells perform other functions at wounds as well, such as clearance of cell and tissue debris during the repair process. However, overenthusiastic inflammation can be damaging[1], so understanding how immune cells are recruited to a wound and how the inflammatory response resolves is crucial. A complex series of temporally overlapping signalling molecules, including the well-studied chemokine proteins, are known to attract immune cells to a wound site. But the identity of the initial chemoattractant generated at the site of tissue damage has, so far, remained elusive. A study published in this issue by Niethammer *et al.*[2] (page 996) now indicates that this earliest 'danger' signal originates from a surprising source — a gradient of hydrogen peroxide (H_2O_2) emanating from the wound.

Neutrophils are the first white blood cells to arrive at a wound site. If blood vessels are damaged by wounding, these cells spill out of the circulation and form part of the clot. In addition, neutrophils migrate to the wound from small blood vessels near the site of tissue damage. To do this, they adhere to the luminal surface of activated endothelial cells lining the blood vessels and then transmigrate either between, or through, endothelial cells into the extravascular space[3]. Subsequently, they move towards the wound by extending projections (pseudopodia) from the cell and retaining those pseudopodia that by chance are directed towards the wound signal[4]. The first cells arrive within minutes of tissue damage: live imaging of fluorescently tagged neutrophils in mice[5] shows that about a million cells have arrived at a small skin lesion 4 hours after the initial tissue insult, and after 18 hours about 5 million have arrived. If the wound is infected, then at least twice this number of neutrophils will be recruited.

The initial signal that instigates rapid neutrophil influx to a wound must fulfil certain criteria: it cannot feasibly be upregulated by increased gene transcription, or by increased protein production through messenger RNA translation, because such forms of modulation are too slow. In line with this, several potential candidates have been mooted as first danger signals, including ATP[6], uric acid[7], the chromatin protein HMGB1 (ref. 8) and various

growth factors[9] — all of which can be released immediately from ruptured cells at the wound site, and each of which has been shown to have some capacity for chemoattraction in tissue-culture experiments. But Niethammer *et al.*[2] provide convincing evidence that, at least in wounds in wet epithelia, the first danger signal is the reactive oxygen species (ROS) H_2O_2.

The authors study a relatively new model in the immunity field, the zebrafish larva, which because of its translucency and the availability of several lines with fluorescently tagged neutrophils[10–12] offers unrivalled opportunities for imaging the dynamic behaviour of immune cells (Fig. 1, overleaf). Compared with the millions of neutrophils drawn to even the smallest nick in humans, only 20 to 30 cells are drawn to a wound in the larval zebrafish fin — and each of these cells can be tracked with fine spatial and temporal precision.

Niethammer and colleagues[2] use a genetically encoded, fluorescent intracellular H_2O_2 sensor, HyPer, to reveal the wound-triggered gradient of H_2O_2. The gradient initiates at the wound margin, extends 200 micrometres from the wound — far enough to reach the nearest vessels — and is established within 5 minutes of wounding, just preceding the movement of the first neutrophils towards the wound. Although it is often presumed that there are spatial gradients of chemoattractants in tissues, they are exceedingly difficult to visualize, and so watching the gradient as it becomes established and then dissipates with a spatial and temporal profile that befits its postulated function is a rare treat. (See Movie 1 in Supplementary Information to ref. 2.)

The authors use a combination of genetic and pharmacological experiments to identify possible sources of the H_2O_2 signal. The most likely physiological source is one of the NADPH oxidase enzymes that generate ROS during the course of standard metabolic processes[13]. There are several such enzymes in zebrafish larval tissue, and bathing the larvae in drugs that inhibit all NADPH oxidases both prevents the establishment of the H_2O_2 gradient and markedly inhibits neutrophil migration to the wound. To pinpoint the specific NADPH oxidase that generates H_2O_2 at the wound margin, Niethammer *et al.*[2] perform a series of knockdown experiments using small molecules called morpholinos, which prevent the translation of specific mRNAs into proteins. They find that knocking down mRNA translation of a specific NADPH oxidase, Duox,
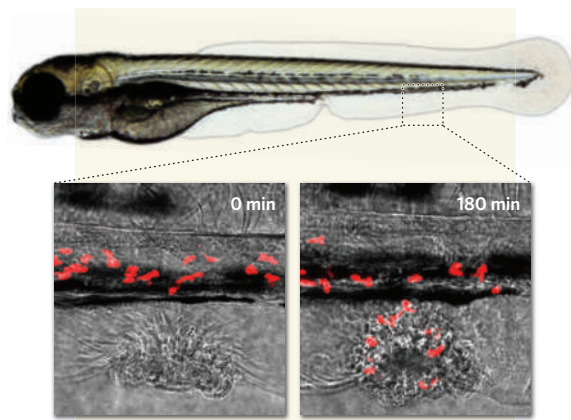
Figure 1 | Visualizing neutrophil migration in zebrafish larvae. A small wound made in the ventral tail fin of a 4-day-old zebrafish larva results in recruitment of fluorescently tagged neutrophils (red) to the wound site within minutes, where they persist for several hours, providing a miniature model of the human wound inflammatory response. Hydrogen peroxide is implicated as the earliest signal recruiting neutrophils to the wound[2].

which is expressed in the larval-fin epithelium, mimics the drug block. There are similar NADPH oxidases in wet epithelia in other organisms — including human gut epithelium — which may well have a similar function in these tissues.

Inflammatory cells themselves are known to release ROS to kill invading microbes and, indeed, cells in the proximity of the wound upregulate several ROS-detoxifying enzymes to protect themselves from these compounds[13]. What an interesting quirk of evolution that the same signal that draws in the first white blood cells to a wound might also double up as an early chemical sterilizer of the wound and adjacent tissues.

There are, of course, still many unanswered questions about the role of H_2O_2 as a chemo-attractant. How does wounding trigger activation of the Duox enzyme to generate H_2O_2? Is H_2O_2 generated by the damaged cells or by their neighbours, or is it produced in response to another as yet unidentified signal that rapidly diffuses from the wound? And how do neutrophils sense the H_2O_2 gradient — do they carry surface receptors that interact with H_2O_2? Or do they have intracellular H_2O_2 sensors such as phosphatase enzymes that could fulfil this role? Indeed, one such phosphatase, PTEN, which is known to have a pivotal role in cell migration, is modulated by exposure to H_2O_2 (ref. 14).

There are no time-points later than about 1 hour in Niethammer and colleagues' study[2], but it will be interesting to discover whether blocking the early H_2O_2 signal shuts the door completely on a wound inflammatory response, or merely delays or dampens it. It will also be crucial to find out if H_2O_2 is a central player in early wound inflammation in other models, and whether other NADPH oxidases perform similar functions in dry epithelia. Further goals will be to identify the downstream signalling molecules and the cues that encourage immune cells to leave the wound site once their work is done. The zebrafish promises to offer more answers to several of these puzzles in the near future. ■

Paul Martin and Yi Feng are in the Departments of Physiology and Pharmacology, and of Biochemistry, University of Bristol, Bristol BS8 1TD, UK.
e-mails: paul.martin@bristol.ac.uk;
yi.feng@bristol.ac.uk

1. Stramer, B. M., Mori, R. & Martin, P. J. Invest. Dermatol. **127,** 1009–1017 (2007).
2. Niethammer, P., Grabher, C., Look, A. T. & Mitchison, T. J. Nature **459,** 996–999 (2009).
3. Vestweber, D. Immunol. Rev. **218,** 178–196 (2007).
4. Cvejic, A. et al. J. Cell Sci. **121,** 3196–3206 (2008).
5. Kim, M.-H. et al. J. Invest. Dermatol. **128,** 1812–1820 (2008).
6. Linden, J. Science **314,** 1689–1690 (2006).
7. Shi, Y., Evans, J. E. & Rock, K. L. Nature **425,** 516–521 (2003).
8. Scaffidi, P., Misteli, T. & Bianchi, M. E. Nature **418,** 191–195 (2002).
9. McNeil, P. L. & Ito, S. J. Cell Sci. **96,** 549–556 (1990).
10. Mathias, J. R. et al. J. Leukocyte Biol. **80,** 1281–1288 (2006).
11. Renshaw, S. A. et al. Blood **108,** 3976–3978 (2006).
12. Hall, C. et al. BMC Dev. Biol. **7,** 42 (2007).
13. Schäfer, M. & Werner, S. Pharmacol. Res. **58,** 165–171 (2008).
14. Lee, S.-R. et al. J. Biol. Chem. **277,** 20336–20342 (2002).

CONDENSED-MATTER PHYSICS

Coupled vibrations

Pertti J. Hakonen and Mika A. Sillanpää

Demonstrating that macroscopic objects can display quantum behaviour, which is usually associated with the microscopic world of atoms, is a long-standing goal in physics. That goal is now within closer reach.

In this issue, LaHaye and colleagues[1] take an important step towards the observation of quantum phenomena in nearly macroscopic moving objects. On page 960, they report experimental evidence of an intriguing interplay between a superconducting artificial atom[2] and a micrometre-size mechanical resonator. Remarkably, their findings can be described using the 'language' of radiation–matter interactions, which has also been successful in explaining the coupling of a superconducting artificial atom to microwave photons[3].

When physicists study matter or energy, they notice that things tend to appear in discrete packets, called quanta. A familiar example of quanta are photons, which are the carriers of electromagnetic energy and the basic constituents of visible light. This 'quantized' nature of matter seems to be pertinent only at the scales of single atoms or molecules. But the laws of physics do not rule out quanta — or indeed any related phenomenon predicted by quantum mechanics — on macroscopic scales. What often happens is that the 'quantum-ness' is washed out by disturbances from the surrounding world.

At sufficiently low temperatures, however, such environmental noise can become small enough for sizeable objects to sustain quantum features. A well-known example is that of superconductivity. In a superconducting wire, a gap of a few nanometres constitutes a Josephson junction. As the temperature is lowered, electron waves consisting of up to billions of individual electrons interfere across the gap and give rise to nearly macroscopic, quantized energy levels. A Josephson junction in such a state can be thought of as an artificial atom, because it displays phenomena analogous to those of atomic physics[2]. For example, researchers have shown how such a junction can interact with an electromagnetic (microwave) resonator[3] in a manner similar to the electromagnetic field in cavity-quantum-electrodynamics systems, in which atoms and light interact in a cavity.

But whereas a vast spectrum of quantum phenomena has been observed in Josephson junctions, observation of corresponding phenomena in bulky objects is lacking. Ultimately, researchers hope to observe quantum behaviour on scales that approach the macroscopic scales of everyday life. One class of model system that offers the potential to achieve that goal is moving bodies, which are often in the form of vibrating strings or bars. However, demonstrating that these systems can exhibit quantum behaviour is extremely hard because of the smallness of the energy quanta of their mechanical vibrations, which translates into extremely small vibrational motions.

LaHaye and colleagues[1] study the vibrations of a micromechanical resonator by coupling it to an artificial atom. Their artificial atom is a charge quantum bit (or qubit)[2], a two-state system which consists of a micrometre-size, superconducting island confined between two Josephson junctions; the two states of the qubit are the charged and uncharged modes of
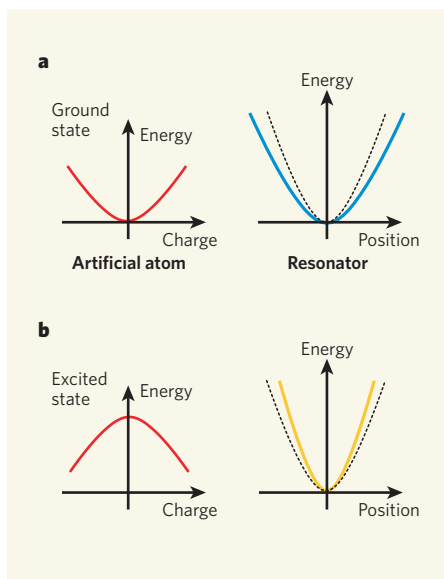
Figure 1 | Coupling a resonator to an artificial atom. LaHaye and colleagues[1] have coupled a micrometre-size mechanical resonator to an artificial atom. The artificial atom is a superconducting charge quantum bit (qubit) — a two-state system whose states depend on whether the qubit is charged or uncharged; the resonator is a suspended silicon nitride beam clamped at both ends. The authors show that the vibrational motion of the resonator induces a variation of the qubit charge and, as a consequence, alters the oscillation frequency of the resonator. **a**, In the ground state, the artificial atom behaves like an electrical capacitor, drawing charge away from the resonator and storing energy. This, in turn, tends to slow down the resonant frequency of the resonator: the curvature of the line (blue) that depicts the resonator's energy as a function of the resonator's position, the beam deflection, is smaller than that of the free (non-coupled) resonator (dashed line). **b**, In the excited state, the artificial atom acts like an energy source[9,10] and the resonant frequency is enhanced (yellow line).

the island. By using a sophisticated lithography technique, the authors fabricated a suspended silicon nitride beam — the micromechanical resonator — clamped at both ends. Silicon nitride is a stiff, low-dissipation material that yields a characteristic resonant frequency of about 60 megahertz for the resonator. The beam was brought into close proximity (as close as 300 nanometres) with the qubit to enable the electrical capacitance (the ability of a device to store charge) between it and the resonator to be as large as possible. Such proximity is crucial for the success of the experiment because it governs the coupling between the qubit and the resonator vibrations, the phonons.

To probe the coupling of the resonator–qubit system, LaHaye and colleagues measured the frequency of the beam vibrations. To do this, they placed the resonator into an electrical circuit that oscillates at the same frequency as the resonator. The circuit is thus used as a means of reading out the resonator's mechanical vibrations electrically[4]. To increase the coupling capacitance, which determines the coupling energy between the qubit and the resonator, a fixed voltage was applied to the beam. Depending on the instantaneous position of the resonator with respect to the qubit island, the beam draws charge — the amount of which is proportional to its voltage and the coupling capacitance — away from the Josephson junctions and alters the charge of the qubit. This amount of charge is related to the qubit state, and is detected as a small change in the frequency of the resonator (Fig. 1).

But without a clear point of comparison, such a small frequency shift — no more than a few kilohertz in 60 megahertz — would easily go unnoticed. Yet the authors were able to detect this shift and control the resonator's frequency by applying an external magnetic field and voltage to the qubit island. These control parameters were used to modulate the properties of the qubit periodically, and the ensuing frequency shifts were found to match

exactly the theoretical predictions[5]. LaHaye and colleagues' findings demonstrate that the resonator phonons behave in much the same way as radiation does in cavity-quantum-electrodynamics systems, allowing a variety of quantum-optics experiments to be performed with mechanical systems.

From a theoretical standpoint, the work of LaHaye *et al.*[1] implies that it should also be possible to discern the individual energy quanta, or other quantum-mechanical behaviour, such as the beam being in a superposition state — one in which the beam is in 'two different

places at the same time'. But pinpointing such fragile quantum features will be challenging for experimentalists. The resonator's frequency will need to be increased by almost an order of magnitude to eliminate thermal disturbances to the number of vibrational quanta and hence to isolate the quantum signal. Various other issues — such as low-frequency charge noise caused by changes in the distribution of atomic bonds within the qubit, or the degradation of the qubit's superconducting properties by non-superconducting single electrons that enter the island intermittently — could also hinder future progress. That said, LaHaye and colleagues' work, along with progress in the field of cavity micromechanics[6–8], suggests that the quantumness of macroscopic moving objects could be detected in laboratory experiments within the next decade.   ∎

Pertti J. Hakonen and Mika A. Sillanpää are at the Low Temperature Laboratory, Helsinki University of Technology, Helsinki FI-02015 TKK, Finland.
e-mail: pjh@boojum.hut.fi

1. LaHaye, M. D., Suh, J., Echternach, P. M., Schwab, K. C. & Roukes, M. L. *Nature* **459,** 960–964 (2009).
2. Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. *Nature* **398,** 786–788 (1999).
3. Wallraff, A. *et al. Nature* **431,** 162–167 (2004).
4. Truitt, P. A., Hertzberg, J. B., Huang, C. C., Ekinci, K. L. & Schwab, K. C. *Nano Lett.* **7,** 120–126 (2007).
5. Irish, E. K. & Schwab, K. *Phys. Rev. B* **68,** 155311 (2003).
6. Gigan, S. *et al. Nature* **444,** 67–70 (2006).
7. Schliesser, A. *et al. Nature Phys.* **4,** 415–419 (2008).
8. Regal, C. A., Teufel, J. D. & Lehnert, K. W. *Nature Phys.* **4,** 555–560 (2008).
9. Sillanpää, M. A. *et al. Phys. Rev. Lett.* **95,** 206806 (2005).
10. Duty, T. *et al. Phys. Rev. Lett.* **95,** 206807 (2005).

CELL BIOLOGY

Beyond the prion principle

Adriano Aguzzi

It seems that many misfolded proteins can act like prions — spreading disease by imparting their misshapen structure to normal cellular counterparts. But how common are bona fide prions really?

The protein-only hypothesis of prion propagation is steadily gaining ground. First envisaged by John Stanley Griffith[1] and later formalized by Stanley Prusiner[2], this theory proposes the existence of an infectious agent composed solely of protein. Three reports, two in *Nature Cell Biology*[3,4] and one in *The Journal of Cell Biology*[5], now contend that, far from being confined to the rare prion diseases, prion-like transmission of altered proteins may occur in several human diseases of the brain and other organs.

Prions are now accepted as causing the transmissible spongiform encephalopathies, which include scrapie in sheep, bovine spongiform encephalopathy (BSE, or mad cow disease) and its human variant Creutzfeldt–Jakob disease. The infectious prion particle is made

up of PrP^{Sc}, a misfolded and aggregated version of a normal protein known as PrP^{C}. Like the growth of crystals, PrP^{Sc} propagates by recruiting monomeric PrP^{C} into its aggregates — a process that has been replicated *in vitro*[6] and in transgenic mice[7]. The breakage of PrP^{Sc} aggregates represents the actual replicative event, as it multiplies the number of active seeds[8].

Apart from prion diseases, the misfolding and aggregation of proteins into various harmful forms, which are collectively known as amyloid, causes a range of diseases of the nervous system and other organs. The clinical characteristics of amyloidoses, however, gave little reason to suspect a relationship to prion diseases.

Hints of prion-like behaviour in amyloid have emerged from studies of Alzheimer's disease

and Parkinson's disease. Alzheimer's disease had been suspected to be transmissible for some time: an early report[9] of disease transmission to hamsters through white blood cells from people with Alzheimer's disease caused great consternation, but was never reproduced. Much more tantalizing evidence came from the discovery[10,11] that aggregates of the amyloid-β (Aβ) peptide found in the brain of people with Alzheimer's disease could be transmitted to the brain of mice engineered to produce large amounts of the Aβ precursor protein APP. Another study[12] has shown that healthy tissue grafted into the brain of people with Parkinson's disease acquires intracellular Lewy bodies — aggregates of the Parkinson's disease-associated protein α-synuclein. This suggests prion-like transmission of diseased protein from the recipient's brain to the grafted cells.

These findings[10–12] raise a provocative question. If protein aggregation depends on the introduction of 'seeds' and on the availability of the monomeric precursor, and if, as has been suggested[13], amyloid represents the primordial state of all proteins, wouldn't all proteins — under appropriate conditions — behave like prions in the presence of sufficient precursor?

Acceptance of this concept is gaining momentum. For one thing, an increasing wealth of traits is being found in yeast, fungi and bacteria that can best be explained as prion-like phenomena (see table). And now, Ren and colleagues[3] provide evidence for prion-like spread of polyglutamine (polyQ)-containing protein aggregates, which are similar to the aggregates found in Huntington's disease. They show that polyQ aggregates can be taken up from the outside by mammalian cells. Once in the cytosol, the polyQ aggregates can grow by recruiting endogenous polyQ.

Clavaguera et al.[4] report similar findings in a mouse model of tauopathy, a neurodegenerative disease caused by intraneuronal aggregation of the microtubule-associated tau protein. Injection of mutant human tau into the brain of mice overexpressing normal human tau transmitted tauopathy, with intracellular aggregation of previously normal tau and spread of aggregates to neighbouring regions of the brain. Notably, full-blown tauopathy was not induced in mice that did not express human tau. Assuming that tau pathology wasn't elicited by some indirect pathway (some mice overexpressing mutated human tau develop protein tangles even when exposed to unrelated amyloid aggregates[14]), this sequence of events is reminiscent of prions. Finally, Frost and colleagues[5] show that extracellular tau aggregates can be taken up by cells in culture. Hence, tau can attack and penetrate cells from the outside, sporting predatory behaviour akin to that of prions.

Yet there is one crucial difference between actual prion diseases and diseases caused by other prion-like proteins (let's call them prionoids) described so far (see table). The behaviour of prions is entirely comparable to that of any other infectious agent: for instance, prions are transmissible between individuals and often across species, and can be assayed with classic microbiological techniques, including titration by bioassay. Accordingly, prion diseases were long thought to be caused by viruses, and BSE created a worldwide panic similar to that currently being provoked by influenza. By contrast, although prionoids can 'infect' neighbouring molecules and sometimes even neighbouring cells, they do not spread within communities or cause epidemics such as those seen with BSE.

So, should any amyloid deserve an upgrade to a bone fide prion status? Currently, amyloid A (AA) amyloidosis may be the most promising candidate for a truly infectious disease caused by a self-propagating protein other than PrP[Sc]. AA amyloid consists of orderly aggregated fragments of the SAA protein, and its deposition damages many organs of the body. Seeds of AA amyloid can be excreted in faeces[15], and can induce amyloidosis if taken up orally (at least in geese)[16]. Also, AA amyloid may be transmitted between mice by transfusion of white blood cells[17]. So, like enteroviruses and, perhaps, sheep scrapie prions, AA amyloid seems to display all the elements of a complete infectious life cycle, including uptake, replication and release from its host.

There are intriguing evolutionary implications to the above findings. If prionoids are ubiquitous, why didn't evolution erect barriers to their pervasiveness? Maybe it is because the molecular transmissibility of aggregated states can sometimes be useful. Indeed, aggregation of the Sup35 protein, which leads to a prion-like phenomenon in yeast, may promote evolutionary adaptation by allowing yeast cells to temporarily activate DNA sequences that are normally untranslated[18]. Mammals have developed receptors for aggregates, and ironically PrP[C] may be one of them[19], although these receptors have not been reported to mediate protective functions. Therefore, we shouldn't be shocked if instances of beneficial prionoids emerge in mammals as well. ∎

Adriano Aguzzi is at the Institute of Neuropathology, University Hospital of Zurich, CH-8091 Zurich, Switzerland.
e-mail: adriano.aguzzi@usz.ch

1. Griffith, J. S. Nature 215, 1043–1044 (1967).
2. Prusiner, S. B. Science 216, 136–144 (1982).
3. Ren, P.-H. et al. Nature Cell Biol. 11, 219–225 (2009).
4. Clavaguera, F. et al. Nature Cell Biol. doi:10.1038/ncb1901 (2009).
5. Frost, B., Jacks, R. L. & Diamond, M. I. J. Biol. Chem. 284, 12845–12852 (2009).
6. Castilla, J., Saá, P., Hetz, C. & Soto, C. Cell 121, 195–206 (2005).
7. Sigurdson, C. J. et al. Proc. Natl Acad. Sci. USA 106, 304–309 (2009).
8. Aguzzi, A. & Polymenidou, M. Cell 116, 313–327 (2004).
9. Manuelidis, E. E. et al. Proc. Natl Acad. Sci. USA 85, 4898–4901 (1988).
10. Kane, M. D. et al. J. Neurosci. 20, 3606–3611 (2000).
11. Meyer-Luehmann, M. et al. Science 313, 1781–1784 (2006).
12. Li, J.-Y. et al. Nature Med. 14, 501–503 (2008).
13. Chiti, F. & Dobson, C. M. Annu. Rev. Biochem. 75, 333–366 (2006).
14. Götz, J., Chen, F., van Dorpe, J. & Nitsch, R. M. Science 293, 1491–1495 (2001).
15. Zhang, B. et al. Proc. Natl Acad. Sci. USA 105, 7263–7268 (2008).
16. Solomon, A. et al. Proc. Natl Acad. Sci. USA 104, 10998–11001 (2007).
17. Sponarova, J., Nyström, S. N. & Westermark, G. T. PLoS ONE 3, e3308 (2008).
18. True, H. L. & Lindquist, S. L. Nature 407, 477–483 (2000).
19. Laurén, J. et al. Nature 457, 1128–1132 (2009).

PRIONS AND POTENTIAL PRIONOIDS

Disease	Protein	Molecular transmissibility	Infectious life cycle
Prion diseases	PrP[Sc]	Yes	Yes
Alzheimer's disease	Amyloid-β	Yes	Not shown
Tauopathies	Tau	Yes	Not shown
Parkinson's disease	α-Synuclein	Host-to-graft	Not shown
AA amyloidosis	Amyloid A	Yes	Possible
Huntington's disease	Polyglutamine	Yes	Not shown

Phenotype	Protein	Molecular transmissibility	Infectious life cycle
Suppressed translational termination (yeast)	Sup35	Yes	Not shown
Heterokaryon incompatibility (filamentous fungi)	Het-s	Yes	Not shown
Biofilm promotion (bacteria)	CsgA	Yes	Not shown

In humans and animals, infectious prion diseases are caused by PrP[Sc], which spreads by recruiting its monomeric precursor PrP[C] into aggregates. Aggregates then multiply by breakage, a process that is termed molecular transmissibility. Other proteins involved in disease and in phenotypes of fungi and bacteria, can also undergo self-sustaining aggregation, but none of these 'prionoid' proteins behaves like typical infectious agents, nor do any of them enact a complete infectious life cycle — with the possible exception of AA amyloid.

FEATURE

Unlocking the secrets of the genome

Despite the successes of genomics, little is known about how genetic information produces complex organisms. A look at the crucial functional elements of fly and worm genomes could change that.

Susan E. Celniker, Laura A. L. Dillon, Mark B. Gerstein, Kristin C. Gunsalus, Steven Henikoff, Gary H. Karpen, Manolis Kellis, Eric C. Lai, Jason D. Lieb, David M. MacAlpine, Gos Micklem, Fabio Piano, Michael Snyder, Lincoln Stein, Kevin P. White and Robert H. Waterston, for the modENCODE Consortium

The primary objective of the Human Genome Project was to produce high-quality sequences not just for the human genome but also for those of the chief model organisms: *Escherichia coli*, yeast (*Saccharomyces cerevisiae*), worm (*Caenorhabditis elegans*), fly (*Drosophila melanogaster*) and mouse (*Mus musculus*). Free access to the resultant data has prompted much biological research, including development of a map of common human genetic variants (the International HapMap Project)[1], expression profiling of healthy and diseased cells[2] and in-depth studies of many individual genes. These genome sequences have enabled researchers to carry out genetic and functional genomic studies not previously possible, revealing new biological insights with broad relevance across the animal kingdom[3,4].

Nevertheless, our understanding of how the information encoded in a genome can produce a complex multicellular organism remains far from complete. To interpret the genome accurately requires a complete list of functionally important elements and a description of their dynamic activities over time and across different cell types. As well as genes for proteins and non-coding RNAs, functionally important elements include regulatory sequences that direct essential functions such as gene expression, DNA replication and chromosome inheritance.

Although geneticists have been quick to decode the functional elements in the yeast *S. cerevisiae,* with its small compact genome and powerful experimental tools[5–6], our understanding of the more complex genomes of human, mouse, fly and worm is still rudimentary. Intrinsic signals that define the boundaries of protein-coding genes can only be partly recognized by current algorithms, and signals for other functional elements are even harder to find and interpret. Experimental approaches, notably the sequencing of complementary DNA and expressed sequence tags, have been invaluable, but unfortunately these data sets remain incomplete[7]. Non-coding RNA genes present an even greater challenge[8–10], and many remain to be discovered, particularly those that have not been strongly conserved during evolution. Flies and worms have roughly the same number of known transcription factors as humans[11], but comprehensive molecular studies of gene regulatory networks have yet to be tackled in any of these species.

In an attempt to remedy this situation, the National Human Genome Research Institute (NHGRI) launched the ENCODE (Encyclopedia of DNA Elements) project in 2003, with the goal of defining the functional elements in the human genome. The pilot phase of the project focused on 1% of the human genome and a parallel effort to foster technology development[12]. The initial ENCODE analysis revealed new findings but also made clear just how complex the biology is and how our grasp of it is far from complete[13]. On the basis of this experience, the NHGRI launched two complementary programmes in 2007: an expansion of the human ENCODE project to the whole genome (www.genome.gov/ENCODE) and the model organism ENCODE (modENCODE) project to generate a comprehensive annotation of the functional elements in the *C. elegans* and *D. melanogaster* genomes (www.modencode.org; www.genome.gov/modENCODE).

These two model organisms, with their ease of husbandry and genetic manipulation, are pillars of modern biological research, and a systematic catalogue of their functional genomic elements promises to pave the way to a more complete understanding of the human genome. Studies of these animals have provided key insights into many basic metazoan processes, including developmental patterning, cellular signalling, DNA replication and inheritance, programmed cell death and RNA interference (RNAi). The genomes are small enough to be investigated comprehensively with current technologies and findings can be validated *in vivo*. The research communities that study these two organisms will rapidly make use of the modENCODE results, deploying powerful experimental approaches that are often not possible or practical in mammals, including genetic, genomic, transgenic, biochemical and RNAi assays. modENCODE, with its potential for biological validation, will add value to the human ENCODE effort by illuminating the relationship between molecular and biological events.

The modENCODE project (Table 1) complements other systematic investigations into these highly studied organisms. In both organisms, RNAi collections have been developed and used to uncover novel gene functions[14–18]. Mutants are being recovered through insertional mutagenesis[19] and targeted deletions (http://celeganskoconsortium.omrf.org;

TABLE 1 modENCODE CONSORTIUM

| Elements	Worm	Fly	Primary experimental data
Transcripts (mRNAs, non-coding RNAs, transcription start sites, untranslated regions, miRNAs)	Robert Waterston (University of Washington), Fabio Piano (New York University)	Susan Celniker (LBNL), Eric Lai (Sloan-Kettering Institute)	Tiling arrays, RNASeq, RT-PCR/RACE, mass spectrometry, 3′ untranslated region clone library, UAS-miRNA flies, knockdowns of RNA-binding proteins
Transcription-factor-binding sites	Michael Snyder (Yale University)	Kevin White (University of Chicago)	ChIP-chip, ChIP-seq, transcription-factor-tagged strains, anti-transcription factor antibodies
Chromatin marks	Jason Lieb (University of North Carolina), Steven Henikoff (University of Washington)	Gary Karpen (LBNL), Steven Henikoff	ChIP-chip and ChIP-seq of chromosome-associated proteins and nucleosomes
DNA replication		David MacAlpine (Duke University Medical Center)	ChIP-chip and ChIP-seq of essential initiator proteins, origin mapping and DNA copy number in differentiated tissues
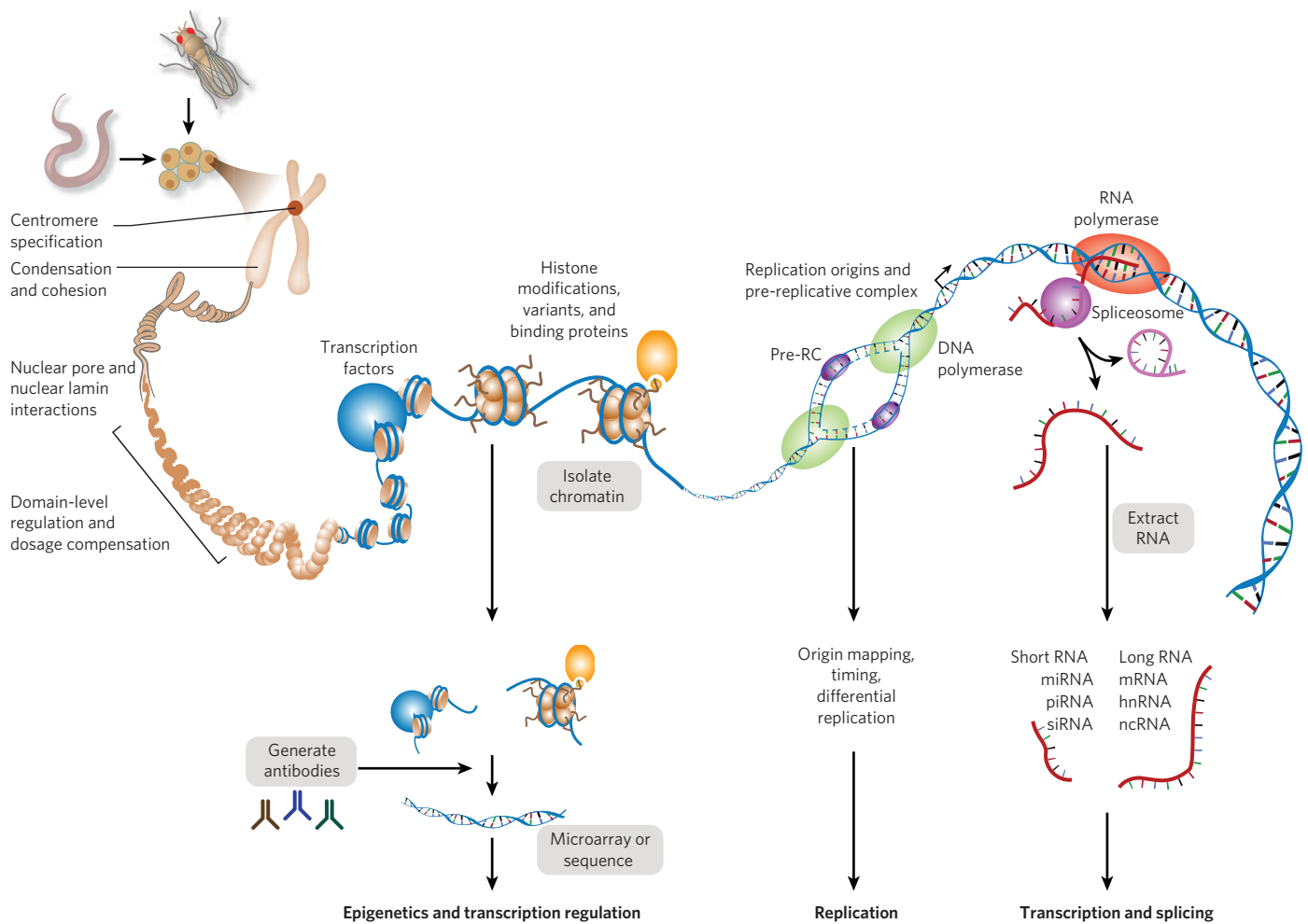
Figure 1 | DNA element functions and identification process.

www.shigen.nig.ac.jp/c.elegans), with the eventual goal of one for every known gene. Genome sequences of related species are now also available for both fly[20,21] and worm[22], and multiple independent wild isolates are being characterized (T. MacKay, personal communication, www.dpgp.org[23]; R.H.W.). First-generation catalogues have been assembled of gene expression patterns during development and in different tissues[24–34].

Research and analysis

The modENCODE project will operate as an open consortium and participants can join on the understanding that they will abide by the set criteria (www.genome.gov/26524644). An important aim of the project is to respond to the needs of the broader *Drosophila* and *C. elegans* scientific communities, and several avenues will be open for suggestions on which experiments to prioritize. For example, researchers can visit www.modencode.org/Vote.shtml now to help prioritize transcription factors for studies using chromatin immuno-precipitation followed by DNA microarray or DNA sequencing (ChIP-chip and ChIP-seq), and can also indicate whether they have useful antibodies. We will seek community input on

other issues as the opportunities arise.

The core of the modENCODE project consists of ten groups who use high-throughput methods to identify functional elements (see Table 1). A Data Coordinating Center (DCC) will collect, integrate and display the data. Together, the groups expect to identify the principal classes of functional element for *D. melanogaster* and *C. elegans*. They will work closely together to complete the precise annotation of protein-coding genes, identify small RNAs and non-coding RNA transcripts, map transcription start sites, identify promoter motif elements, elucidate functional elements within 3′ untranslated regions, and identify alternatively spliced transcripts as well as the signals required for splicing. Genomic sites bound by sequence-specific transcription factors will also be comprehensively identified. Charting the chromatin 'landscapes' will include the characterization of key histone modifications and variants, nucleosome phasing, RNA polymerase II isoforms and proteins involved in dosage compensation, centromere function, replication, homologue pairing, recombination and associations of chromosomes with the nuclear envelope.

Integrative analysis of these data across

the different types of functional element will be used to reveal fundamental principles of fly and worm genome biology and to begin to uncover the emergent properties of these complex genomes. Some topics the modENCODE groups, along with interested members of the wider community, intend to explore are outlined below, but these are only a beginning. Our intention is to create a resource that will provide the foundation for ongoing analysis by scientists for years to come.

Our two model organisms share many similarities with other metazoans, including humans. They also differ from other organisms in some striking ways, particularly in details of the establishment and maintenance of cellular identity, centromere biology and heterochromatin function. To help understand how the similarities and differences in worm and fly biology are reflected in their genome sequences and how they are specified by genome function at the molecular level, we will carry out comparative analyses of transcription, splicing, *cis*-regulatory and post-transcriptional elements and chromatin function. We will subsequently investigate how our findings apply to the control of gene expression in the human genome.

We also plan to use genome-wide data on pre- and post-transcriptional functional elements to expand our understanding of gene-regulatory networks. We will study how these two layers of control complement or reinforce each other during development. For example, the availability of full-length transcripts and promoter structures for microRNA (miRNA) genes will enable us to develop models of regulatory circuits that integrate the upstream regulation of miRNA genes with that of other regulatory factors (such as transcription factors) and the effects of miRNAs on their downstream targets. We will search global patterns identified in the regulatory programs for emerging principles of gene regulation within and across species; as part of this endeavour, we will evaluate evidence for the modular structure of regulatory networks.

Because several developmental stages and diverse tissues will be sampled in both animals, we will be able to investigate the global and dynamic activities of functional elements across the entire genome in multiple cell types and stages of differentiation. We aim to define the characteristics and rules that distinguish regulatory programs in different cell types and developmental stages at the DNA, chromatin, and post-transcriptional levels. This will enable us to identify the types of element that function together in various spatio-temporal environments and find new types of functional element, perhaps including those used in restricted developmental contexts.

An important objective is to generate specific biological hypotheses that can be refined and tested experimentally by the broader scientific community. For example, these analyses might identify transcribed regions with novel regulatory roles, structural regions that function in the establishment of chromatin structure or three-dimensional conformation, enhancers far away from the gene they control, and alternative promoter regions. In addition, we will use comparative analyses of the sequenced genomes from different species to clarify the extent of conservation and the functional constraints associated with potential new classes of element and to characterize their evolutionary signatures[21].

Another objective of the modENCODE project is the creation of reference data sets of maximum utility. We have agreed that, whenever possible, a common set of reagents will be used to facilitate comparison of data sets generated by different groups. For example, the fly and worm groups using ChIP-chip and related methods to map the genome-wide distributions of histone modifications will use a common set of validated antibodies. In addition, we will use common fly and worm strains, and in the case of *Drosophila*, the common cell lines Kc167, S2-DRSC, CME W1 Cl.8+ and ML-DmBG3-c2.

The fly and worm genomes are about a thirtieth of the size of their mammalian counterparts, making current methods for high-throughput genomic analysis cost-effective. We will use high-density tiling DNA microarrays to interrogate the genome on a single microarray (*C. elegans*, 26 base pair (bp) median spacing; *D. melanogaster*, 38 bp median spacing) at a resolution sufficient for ChIP-chip experiments. Denser arrays (*D. melanogaster*, 7 bp median spacing), which promise higher resolution, will be used in a move to high-throughput sequencing platforms such as the Illumina Genome Analyzer to generate sufficient sequence coverage for transcript mapping and miRNA and ChIP experiments.

The biological significance of the genomic features identified will be tested in experiments designed to evaluate the accuracy and functionality of subsets of the structural and regulatory annotations. For example, we will carry out ChIP experiments on extracts from whole animals or cells that lack selected regulators (using mutants or RNAi). The tissue-specific DNA-binding patterns of selected regulators will be validated in transgenic animals. Figure 1 summarizes the DNA elements to be interrogated and the methods to be used.

Data management and accessibility

Data generated by the modENCODE Consortium, including those from validation experiments, will be collected, quality checked, integrated and distributed through the modENCODE DCC (www.modencode.org). The DCC will collate detailed metadata for each submitted data set to ensure broad and long-term usability. Where appropriate, the data will also be submitted to public databases, for example, GenBank (www.ncbi.nlm.nih.gov) and the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) or Array Express (www.ebi.ac.uk/microarray-as/aer/entry) and the University of California, Santa Cruz Genome Bioinformatics Site (http://genome.ucsc.edu). The DCC will also work closely with WormBase (www.wormbase.org) and FlyBase (www.flybase.org) to facilitate integration of the modENCODE data with selected data from these databases and with other information about these organisms.

All data will be available for bulk download through an FTP site and through a number of Generic Model Organism Database tools (www.gmod.org): BioMart (www.biomart.org) will provide powerful data-mining capabilities, and InterMine (www.intermine.org) will provide a flexible interface for complex querying of the data, a library of canned queries, and powerful list-based tools and operations (http://intermine.modencode.org). As for the ENCODE pilot project data (www.genome.gov/10005107), new data can be examined alongside existing data using interactive genome browsers[35] for both the fly (www.modencode.org/cgi-bin/gbrowse/fly) and the worm (www.modencode.org/cgi-bin/gbrowse/worm).

The *Drosophila* and *C. elegans* communities have thrived because of their open culture. In keeping with this tradition and with those of the genome sequencing projects, HapMap and the ENCODE pilot project, modENCODE is a 'community resource project' subject to the NHGRI's data-sharing policy. The success of this policy is based on mutual and independent responsibilities for the production and use of the resource. We will release data rapidly (Table 2), before publication, once they have been established to be reproducible (verification; see www.modencode.org/'Publication Policy link' for the criteria), even if the data have not been sampled to determine if there is biological meaning (validation). In turn, users are asked to recognize the source of the data and to respect the legitimate interest of the resource producers to publish an initial report of their work (see www.genome.gov/modencode for more details). Finally, the funding agencies

TABLE 2 GLOBAL ANALYSIS GOALS

Elements and processes	Specific examples
Transcribed regions	Define cell- and tissue-specific transcriptional landscapes. Annotate transcription start sites, exons, untranslated region structures, small regulatory RNAs and short single-exon open reading frames
Gene regulation, transcriptional regulation	Identify transcription-factor binding sites in various cell and tissue types. Correlate chromatin structure marks and transcriptional activities for protein-coding and non-protein-coding genes
Post-transcriptional regulation	Identify tissue-specific binding sites for miRNAs and other small RNAs, RNA secondary structures and alternative splicing regulatory motifs
Chromatin structure and function	Identify sites of association between DNA and chromosomal proteins involved in centromere specification, meiotic recombination, dosage compensation, nuclear envelope and matrix interactions and chromosome condensation. Identify sites of incorporation of histone variants and specifically modified histones. Correlate transcription maps for meta-analysis of developmental chromatin dynamics.
DNA replication	Identify cell- and tissue-specific origins of replication. Correlate with cell- and tissue-specific transcription and chromatin marks

recognize the need to support the analysis and dissemination of the data.

In addition, a variety of physical resources (for example, DNA constructs and transgenic strains) will be produced that are likely to be of use to the broader community and to which that community will have unrestricted access. We expect to cooperate with data users in the worm and fly communities to set the gold standard for data release and openness.

Conclusion

The Human Genome Project benefited enormously from the technology developed and the experience acquired in sequencing the significantly smaller genomes of model organisms, particularly *C. elegans* and *D. melanogaster*. The modENCODE project is dedicated to the next phase of decoding the information stored in these genomes: the comprehensive identification of sequence-based functional elements. Having laid the foundation for the discovery of many of the genetic programs underlying metazoan development and behaviour, *Drosophila* and *Caenorhabditis* will serve as ideal model systems to identify DNA-based functional elements on a genome-wide basis. In the future, these data will provide a powerful platform for characterizing the functional networks that direct multicellular biology, thereby linking genomic data with the biological programs of higher organisms, including humans. ■

1. Sabeti, P. C. *et al. Nature* **449,** 913–918 (2007).
2. Neve, R. M. *et al. Cancer Cell* **10,** 515–527 (2006).
3. Chintapalli, V. R., Wang, J. & Dow, J. A. *Nature Genet.* **39,** 715–720 (2007).
4. Nichols, C. D. *Pharmacol. Ther.* **112,** 677–700 (2006).
5. Ross-Macdonald, P. *et al. Nature* **402,** 413–418 (1999).
6. Boone, C., Bussey, H. & Andrews, B. J. *Nature Rev. Genet.* **8,** 437–449 (2007).
7. Celniker, S. E. & Rubin, G. M. *Annu. Rev. Genomics Hum. Genet.* **4,** 89–117 (2003).
8. Tupy, J. L. *et al. Proc. Natl Acad. Sci. USA* **102,** 5495–5500 (2005).
9. Ruby, J. G. *et al. Cell* **127,** 1193–1207 (2006).
10. Ruby, J. G. *et al. Genome Res.* **17,** 1850–1864 (2007).
11. Reece-Hoyes, J. S. *et al. Genome Biol.* **6,** R110 (2005).
12. The ENCODE Project Consortium *Science* **306,** 636–640 (2004).
13. Birney, E. *et al. Nature* **447,** 799–816 (2007).
14. Boutros, M. *et al. Science* **303,** 832–835 (2004).
15. Kamath, R. S. *et al. Nature* **421,** 231–237 (2003).
16. Rual, J. F. *et al. Genome Res.* **14,** 2162–2168 (2004).
17. Sonnichsen, B. *et al. Nature* **434,** 462–469 (2005).
18. Dietzl, G. *et al. Nature* **448,** 151–156 (2007).
19. Bellen, H. J. *et al. Genetics* **167,** 761–781 (2004).
20. Clark, A. G. *et al. Nature* **450,** 203–218 (2007).
21. Stark, A. *et al. Nature* **450,** 219–232 (2007).
22. Stein, L. D. *et al. PLoS Biol.* **1,** E45 (2003).
23. Hillier, L. W. *et al. Nature Methods* **5,** 183–188 (2008).
24. Tomancak, P. *et al. Genome Biol.* **3,** research0088.1–0088.14 (2002).
25. Arbeitman, M. N. *et al. Science* **297,** 2270–2275 (2002).
26. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. *Science* **302,** 249–255 (2003).
27. Li, T. R. & White, K. P. *Dev. Cell* **5,** 59–72 (2003).
28. Stolc, V. *et al. Science* **306,** 655–660 (2004).
29. Manak, J. R. *et al. Nature Genet.* **38,** 1151–1158 (2006).
30. Tomancak, P. *et al. Genome Biol.* **8,** R145 (2007).
31. Jiang, M. *et al. Proc. Natl Acad. Sci. USA* **98,** 218–223 (2001).
32. Reinke, V., Gil, I. S., Ward, S. & Kazmer, K. *Development* **131,** 311–323 (2004).
33. Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L. & Hunter, C. P. *Development* **130,** 889–900 (2003).
34. Kim, S. K. *et al. Science* **293,** 2087–2092 (2001).
35. Stein, L. D. *et al. Genome Res.* **12,** 1599–1610 (2002).

Authors

Susan E. Celniker[1], Laura A. L. Dillon[2], Mark B. Gerstein[3,4], Kristin C. Gunsalus[5], Steven Henikoff[6], Gary H. Karpen[7], Manolis Kellis[8,9], Eric C. Lai[10], Jason D. Lieb[11], David M. MacAlpine[12], Gos Micklem[13], Fabio Piano[5], Michael Snyder[14], Lincoln Stein[15], Kevin P. White[16,17], Robert H. Waterston[18]

[1]Department of Genome Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [2]Division of Extramural Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Program in Computational Biology and Bioinformatics, [4]Department of Computer Science and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. [5]Center for Genomics and Systems Biology, New York University, New York, New York 10003, USA. [6]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [7]Department of Genome and Computational Biology, Lawrence Berkeley National Laboratory, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [8]Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02140, USA. [9]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [10]Sloan-Kettering Institute, New York, New York 10065, USA. [11]Department of Biology and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [12]Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. [13]Department of Genetics, University of Cambridge, CB2 3EH, UK, and Cambridge Systems Biology Centre, Tennis Court Road, Cambridge CB2 1QR, UK. [14]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06824, USA. [15]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11542 USA. [16]Institute for Genomics & Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [17]Institute for Genomics & Systems Biology, Argonne National Laboratory, Argonne, Illinois 60439, USA. [18]Department of Genome Sciences and University of Washington School of Medicine, Seattle, Washington 98195, USA.

REVIEWS

Emergence and pandemic potential of swine-origin H1N1 influenza virus

Gabriele Neumann[1], Takeshi Noda[2] & Yoshihiro Kawaoka[1,2,3,4]

Influenza viruses cause annual epidemics and occasional pandemics that have claimed the lives of millions. The emergence of new strains will continue to pose challenges to public health and the scientific communities. A prime example is the recent emergence of swine-origin H1N1 viruses that have transmitted to and spread among humans, resulting in outbreaks internationally. Efforts to control these outbreaks and real-time monitoring of the evolution of this virus should provide us with invaluable information to direct infectious disease control programmes and to improve understanding of the factors that determine viral pathogenicity and/or transmissibility.

The genomes of influenza viruses are plastic, owing to point mutations and reassortment events that contribute to the emergence of new variants or strains with epidemic or pandemic potential. Inasmuch, influenza A viruses have caused several pandemics during the last century, and continue to cause annual epidemics. Both epidemics and pandemics have substantial economic impact owing to the costs of prevention and treatment, work absenteeism, physician visits and excess hospitalizations. Therefore, a detailed understanding of the mechanisms that determine pathogenicity and interspecies transmission, combined with the availability of effective preventative and therapeutic measures, is critical to the control of influenza virus infections.

Influenza A viruses

Influenza A viruses belong to the family Orthomyxoviridae. On the basis of the antigenicity of their haemagglutinin (HA) and neuraminidase (NA) molecules, they are classified into 16 HA subtypes (H1–H16) and 9 NA subtypes (N1–N9). Influenza A viruses contain a genome composed of eight segments of single-stranded, negative-sense RNA that each encodes one or two proteins (Fig. 1). The HA protein is critical for binding to cellular receptors and fusion of the viral and endosomal membranes (Fig. 2). Replication and transcription of viral RNAs (vRNAs) are carried out by the three polymerase subunits PB2, PB1 and PA, and the nucleoprotein NP. Newly synthesized viral ribonucleoprotein (vRNP) complexes are exported from the nucleus to the cytoplasm by the nuclear export protein (NEP, formerly called NS2) and the matrix protein M1, and are assembled into virions at the plasma membrane. The NA protein facilitates virus release from infected cells by removing sialic acids from cellular and viral HA and NA proteins. The functions of the ion channel protein M2, the interferon antagonist NS1, and the PB1-F2 protein are discussed in more detail later.

Influenza pandemics

Influenza A viruses cause recurrent epidemics and global pandemics. Pandemics are typically caused by the introduction of a virus with an HA subtype that is new to human populations. Two mechanisms that are not mutually exclusive, reassortment and interspecies transmission, result in the introduction of viruses with new HA subtypes into human populations (Fig. 3). The recently emerged swine-origin

H1N1 influenza viruses (S-OIVs) are being detected in an increasing number of countries, and their global spread would undoubtedly result in a considerable number of infected individuals. At present, the mortality rate associated with S-OIV infections seems to be comparable to that of seasonal influenza virus outbreaks; however, increased surveillance activities could reveal a higher rate.

'Spanish' influenza (H1N1). The pandemic of 1918–1919 killed as many as 50 million people worldwide, and remains unprecedented in its severity. A first, mild wave in the spring of 1918 was replaced by a second wave in September to November of 1918 that resulted in mortality rates of over 2.5%, compared to less than 0.1% typically recorded for influenza outbreaks. A third wave with equally high mortality rates swept around the world in 1919.

The mortality pattern of the 'Spanish' influenza was unusual with high mortality rates for young adults. The atypical mortality pattern observed with the Spanish influenza remains unexplained to this day. In contrast, the morbidity pattern was similar to other pandemics—that is, children under the age of 15 experienced the highest attack rates.

The Spanish influenza virus was restricted to the respiratory tract; lack of systemic infection has also been observed in non-human primates experimentally infected with reconstituted 1918 virus[1]. Most patients died of bacterial pneumonia[2], which may be attributed to the lack of antibiotics in 1918–1919; however, many others died owing to viral pneumonia.

Although the Spanish influenza virus was not isolated during the outbreak in 1918–1919, the genomic sequences of this virus were determined[3,4] and revealed an avian-like H1N1 virus that contains human-like signature amino acids in several proteins. The Spanish influenza virus lacks a multibasic HA cleavage site[4], a hallmark of highly pathogenic avian influenza viruses (see 'Role of HA in viral pathogenicity').

Reverse genetics[5] allowed the re-creation of the Spanish influenza virus[6] and its characterization. The Spanish influenza virus elicits aberrant innate immune responses in mice[7] and in non-human primates[1], a feature that it shares with highly pathogenic H5N1 viruses[8] and that probably contributes to pathogenicity and mortality in humans. Further studies showed that the HA protein[9–11], the replication complex[6,12,13], the NS1 protein[14], and the PB1-F2 protein[15] contributed to its virulence, and that the HA and PB2 proteins were critical to its transmissibility[13].

[1]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53711, USA. [2]International Research Center for Infectious Diseases, [3]Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [4]ERATO Infection-Induced Host Responses Project, Japan Science and Technology Agency, Saitama 332-0012, Japan.
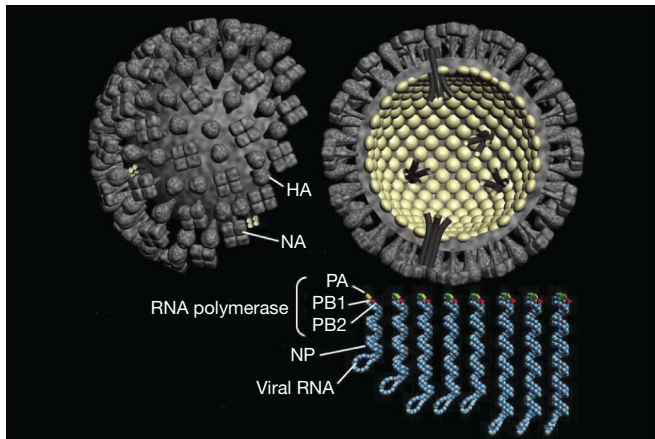
Figure 1 | Schematic diagram of influenza A viruses. Virions are decorated with two surface glycoproteins, HA and NA. The genome is composed of eight segments of single-stranded RNA that interact with the nucleoprotein and the components of polymerase complex (PB2, PB1 and PA).

'Asian' influenza (H2N2). The 'Asian' influenza originated in Southern China in February 1957. From there, it spread to Singapore (March 1957), Hong Kong (April 1957), Japan (May 1957), and the United States and the United Kingdom (October 1957). A second wave was detected in January 1958. In the United States, excess mortality was estimated to be 70,000. The pandemic was caused by a human/avian reassortant that introduced avian virus H2 HA and N2 NA genes into human populations (Fig. 3). Furthermore, the Asian influenza virus also possessed a PB1 gene of avian virus origin.

'Hong Kong' influenza (H3N2). In 1968, viruses of the H2N2 subtype were replaced by another human/avian reassortant that possessed an H3 HA gene of avian virus origin (Fig. 3). Again, the PB1 gene of the pandemic virus was derived from an avian virus. The virus was first isolated in Hong Kong in July 1968 and caused a pandemic in the winters of 1968–1969 and 1969–1970. In the United States, an estimated 33,800 people died from the 'Hong Kong' influenza.

'Russian' influenza (H1N1). In May 1977, an influenza virus outbreak was reported in China that affected young adults in the northern hemisphere in the winter of 1977–1978. The outbreak was caused by influenza viruses of the H1N1 subtype that closely resembled viruses that had circulated in the early 1950s[16], suggesting accidental release of this virus. The re-emerging H1N1 virus did not replace the H3N2 viruses circulating at the time, and both subtypes are co-circulating in humans to this day. Reassortment between viruses of these subtypes resulted in the emergence of H1N2 viruses in human populations in 2001. However, these H1N2 viruses have since disappeared.

Highly pathogenic H5N1 influenza viruses

The infection of 18 individuals in Hong Kong in 1997 with highly pathogenic avian influenza viruses of the H5N1 subtype, which resulted in six fatalities[17,18], marked the first reported fatal infections of humans with avian influenza viruses. This outbreak was brought under control with the depopulation of live birds in poultry markets in Hong Kong. After a period of local and sporadic outbreaks, a new outbreak started in 2003. H5N1 viruses have since reassorted frequently[19–22] and have spread to Europe and Africa and/or become enzootic in poultry populations in many Southeast Asian countries[23].

The highly pathogenic H5N1 viruses have several remarkable features. First, they are not only lethal in chickens, but some highly pathogenic H5N1 viruses also kill waterfowl, the natural reservoir of influenza A viruses. Second, they replicate and cause lethal infection in mice without prior adaptation. Third, they have fatally infected several mammalian species. Non-lethal pig infections have been detected at low rates. Fourth, their pathogenicity in ferrets has increased over the years, indicating the acquisition of mutations that increase pathogenicity in mammalian species. Fifth, and of most concern, is their continued transmission to humans, resulting in severe respiratory infection with high mortality rates.
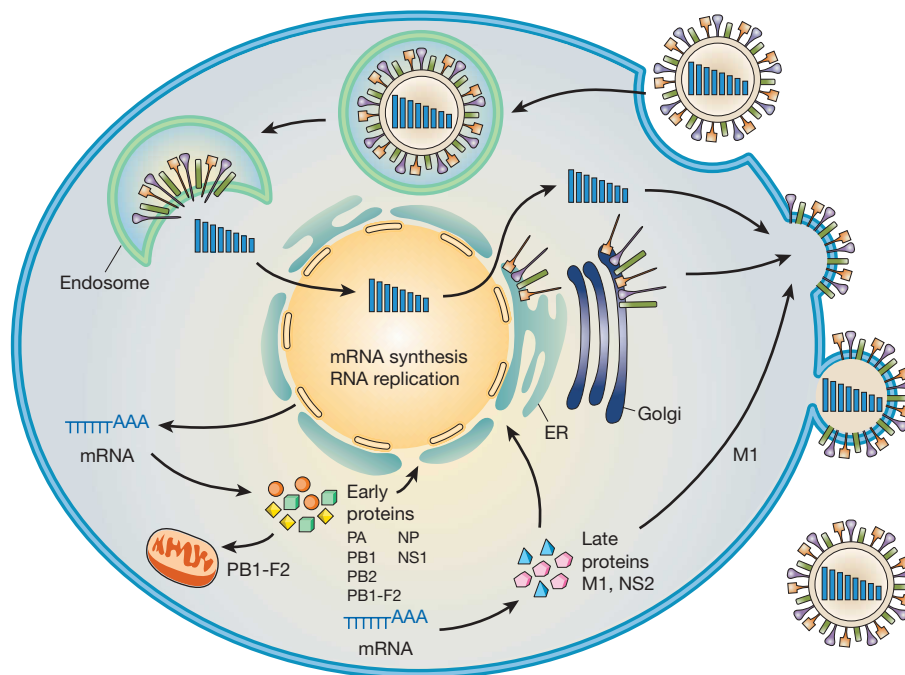
Figure 2 | Schematic diagram of the influenza viral life cycle. After receptor-mediated endocytosis, the viral ribonucleoprotein (vRNP) complexes are released into the cytoplasm and subsequently transported to the nucleus, where replication and transcription take place. Messenger RNAs are exported to the cytoplasm for translation. Early viral proteins, that is, those required for replication and transcription, are transported back to the nucleus. Late in the infection cycle, the M1 and NS2 proteins facilitate the nuclear export of newly synthesized vRNPs. PB1-F2 associates with mitochondria. The assembly and budding of progeny virions occurs at the plasma membrane.
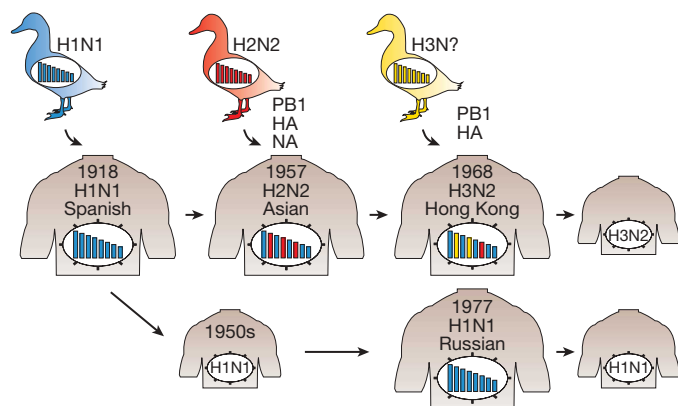
Figure 3 | Emergence of pandemic influenza viruses. The Spanish influenza was probably caused by the transmission of an avian influenza virus to humans. In 1957, the introduction of avian virus H2 HA, N2 NA and PB1 genes into human populations resulted in the Asian influenza. Similarly, the introduction of avian virus H3 HA and PB1 genes into human populations led to the Hong Kong influenza in 1968. In 1977, H1N1 viruses reappeared, which closely resembled strains that had been circulating in the mid-1950s.

As of 15 May 2009, 424 human infections with H5N1 have been confirmed, resulting in 261 deaths (http://www.who.int/csr/disease/avian_influenza/country/cases_table_2009_05_15/en/index.html). Although several family clusters of H5N1 virus infection have been described, sustained human-to-human infection has not occurred. Hence, these H5N1 viruses are characterized by a high mortality rate but inefficient spread among humans; in contrast, S-OIVs seem to spread efficiently among humans but have caused a limited number of fatal infections.

Several reports indicate that highly pathogenic avian H5N1 viruses replicate mainly in the lower respiratory tract of humans[8,24], and that the viral load correlates with the outcome of the infection[8]. Human H5N1 infections cause severe pneumonia and lymphopenia[24–26], and are characterized by high levels of cytokines and chemokines[8,27,28], a finding that was also confirmed in *in vitro* studies[29,30]. The induction of hypercytokinemia and hyperchemokinemia may thus be associated with the level of virus replication[8].

Outbreak of swine-origin H1N1 viruses

Epidemiological data now indicate that an outbreak of influenza-like respiratory illness started in the Mexican town of La Gloria, Veracruz, in mid-February of 2009 (ref. 31; Table 1). In early April, public health authorities in Mexico began investigating high numbers of pneumonia/influenza-like illness, and informed the Pan American Health Organization (PAHO), the regional office of the World Health Organization (WHO), of a possible outbreak. In the United States, the Centers for Disease Control (CDC) identified S-OIVs in two specimens independently collected in Southern California in mid-April. On 23 April, the Public Health Agency of Canada also detected S-OIVs in specimens received from Mexico. Further cases and the finding that the Mexican and Californian cases were caused by similar viruses triggered alerts by the CDC and WHO on 24 April.

By the end of April, international spread and clusters of human-to-human transmission prompted the WHO to increase the pandemic alert from phase 3 to phase 4, and shortly after, to phase 5 (human-to-human spread in at least two countries, and signs of an imminent pandemic). In Mexico, substantial social-distancing measures were implemented. Moreover, massive campaigns were undertaken to educate the public about precautionary hygiene measures. As of 21 May 2009, 41 countries have reported 11,034 cases, including 85 deaths. Most cases outside Mexico and the United States have been caused by travellers from Mexico. Most infections seem to be mild and do not require hospitalization[32]. Careful monitoring will be necessary during the following months (that is, during the winter season in the southern hemisphere) to be prepared for the potential emergence of more virulent variants, as observed with the 1918 pandemic.

Data on the genetic composition of the virus became available soon after viral isolation from the initial cases[32]. The S-OIVs probably resulted from the reassortment of recent North American H3N2 and H1N2 swine viruses (that is, avian/human/swine 'triple' reassortant viruses) with Eurasian avian-like swine viruses[32] (Fig. 4). As a result, these viruses possess PB2 and PA genes of North American avian virus origin, a PB1 gene of human H3N2 virus origin, HA (H1), NP, and NS genes of classical swine virus origin, and NA (N1) and M genes of Eurasian avian-like swine virus origin (hence their original description as 'quadruple' reassortants). However, the human-like PB1 gene and the avian-like PB2 and PA genes have been circulating in pigs since 1997–1998 (when triple reassortant swine viruses were first isolated), and have probably undergone adaptation to pigs. These viruses do not possess markers associated with high pathogenicity (see the following sections on the role of viral proteins in pathogenicity for more details). Unlike negatively stained virions from S-OIVs that appeared spherical (http://www.cdc.gov/h1n1flu/images.htm), our transmission electron microscopic analysis of cells infected with S-OIVs revealed virions of a distinctively filamentous shape (Fig. 5).

Role of HA in viral pathogenicity

Influenza virus pathogenicity is multigenic, and the determinants of pathogenicity may differ among animal species. However, the HA protein has an important role in expressing high pathogenicity in many animal species. It mediates the binding of the virus to host cells, and the subsequent fusion of the viral and endosomal membranes for vRNP release into the cytoplasm. These functions assign a critical role to HA in the viral life cycle.

Receptor distribution on host cells. Influenza virus host specificity can be explained in part by the difference in receptor-binding specificity for human and avian influenza viruses. Human influenza viruses preferentially bind to sialic acid that is linked to galactose by an α2,6-linkage (SAα2,6Gal)[33]. This preference is matched by SAα2,6Gal on epithelial cells in the human trachea. In contrast, avian influenza viruses preferentially recognize SAα2,3Gal that is matched by SAα2,3Gal on epithelial cells in the intestinal tract of waterfowl[34] (the main replication site of avian influenza viruses).

The receptor-binding specificity of human and avian influenza viruses suggests that avian influenza viruses need to acquire the ability

Table 1 | Timeline of swine-origin H1N1 virus outbreak (see also ref. 32)

| Date	Event
Mid-February	Outbreak of respiratory illness in La Gloria, Veracruz, Mexico[31]
12 April	Mexican public health authorities report outbreak in Veracruz to the PAHO
15 April	CDC identifies S-OIVs in the specimen of a boy from San Diego, California
17 April	CDC identifies S-OIVs in the specimen of a girl from Imperial, California
21 April	CDC alerts doctors to a new strain of H1N1 influenza virus
23 April	The Public Health Agency of Canada identifies S-OIVs in specimens from Mexico
24 April	WHO issues Disease Outbreak Notice
27 April	International spread and clusters of human-to-human transmission prompt WHO to raise the pandemic alert from phase 3 to 4
29 April	WHO raises the pandemic alert from phase 4 to 5 (human-to-human spread in at least two countries in one WHO region)
21 May	41 countries report 11,034 cases, including 85 deaths
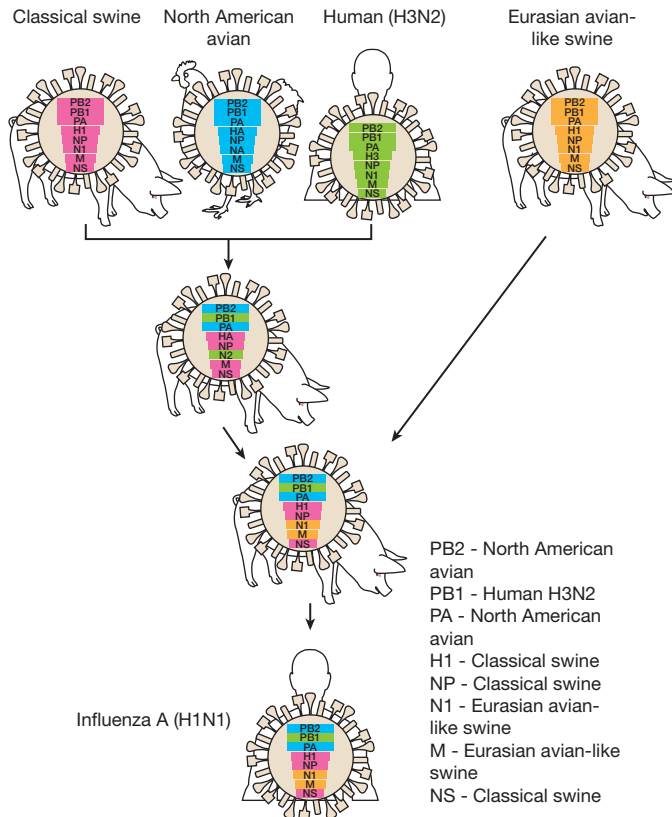
Figure 4 | Genesis of swine-origin H1N1 influenza viruses. In the late 1990s, reassortment between human H3N2, North American avian, and classical swine viruses resulted in triple reassortant H3N2 and H1N2 swine viruses that have since circulated in North American pig populations. A triple reassortant swine virus reassorted with a Eurasian avian-like swine virus, resulting in the S-OIVs that are now circulating in humans.

to recognize human-type receptors to cause a pandemic. Indeed, the earliest isolates of the 1918, 1957 and 1968 pandemics possessed HA that, although of avian origin, recognized human-type receptors. In light of these findings, the infection of humans with highly pathogenic avian H5N1 viruses seemed to be surprising, particularly because the H5N1 viruses isolated from infected individuals in Hong Kong in 1997 preferentially recognized SAα2,3Gal[35]. However, studies showed

Figure 5 | Electron microscopic picture of recently emerged swine-origin H1N1 viruses. Madin-Darby canine kidney cells were infected with A/California/04/09 (H1N1) virus and observed by thin-section electron microscopy 24 h later. Most virus particles showed a filamentous shape of more than 1 μm in length. Scale bar, 1 μm.

avian-type receptors on human epithelial cells that line the respiratory bronchiole and the alveolar walls, but human-type receptors on human epithelial cells in nasal mucosa, paranasal sinuses, pharynx, trachea and bronchi[36,37]. Another study, however, showed the *ex vivo* infection of human upper respiratory organs with an H5N1 avian virus[38]. Still, the finding of avian-type receptors in human lungs explains the severe pneumonia seen in humans with highly pathogenic avian H5N1 viruses.

HA receptor specificity. The differences in receptor-binding specificity of human and avian viruses are determined by the amino acid residues in the HA receptor-binding pocket. Gln at position 226 and Gly at position 228 of H2 and H3 HAs confer binding to avian-type receptors, whereas Leu and Ser at these positions determine binding to human-type receptors. For H1 HAs, amino acids at position 190 and 225 (H3 numbering) determine receptor-binding specificity. HA-Asp 190 and HA-Asp 225 (found in human H1 HAs) confer binding to human-type receptors, whereas HA-Glu 190 and HA-Gly 225 (found in avian H1 HAs) confer binding to avian-type receptors[39]. Two viruses that differ in receptor recognition were circulating during the 1918 pandemic: one recognizing only human-type receptors that transmits efficiently among ferrets, and one recognizing both avian- and human-type receptors that transmits inefficiently in this animal[40]. The recently emerged S-OIVs possess the 'human virus'-type amino acid at positions 190 and 225, probably supporting efficient transmissibility of these viruses in humans. Interestingly, some S-OIV isolates possess an amino acid substitution at position 135 or 226 that has been found in H5N1 viruses isolated from humans and has been shown to affect receptor binding (S. Yamada and Y.K., unpublished data). These mutations may thus reflect viral adaptation in humans, an assumption that needs to be tested.

For H5N1 viruses, amino acid changes at positions 133, 138, 186, 192 and 227 (H3 numbering) have been identified in human isolates and confer human-type receptor recognition[41–43]. Experimental changes at positions 226 and 228 (Gln226Leu and Gly228Ser), but not at position 190 (Glu190Asp), resulted in the recognition of human-type receptors as well as avian-type receptors[44]; however, the respective amino acid changes at positions 226 and 228 have not been detected in human H5N1 virus isolates.

HA cleavage. HA cleavage is essential for viral infectivity because exposure of the amino terminus of HA2 (the so-called 'fusion peptide') mediates fusion between the viral envelope and the endosomal membrane, an essential step for vRNP release to the cytoplasm. HA cleavability is determined by the amino acid sequence at the cleavage site. Low pathogenic avian viruses and non-avian influenza viruses (with the exception of H7N7 equine influenza viruses) possess a single Arg residue at the cleavage site that is cleaved by proteases in the respiratory and/or intestinal organs, and hence restricts viral replication locally. In contrast, highly pathogenic H5 and H7 viruses possess several basic amino acids at the HA cleavage site[45] (Table 2). This motif is recognized by ubiquitous proteases, such as furin and PC6 (also known as PCSK5), and leads to systemic infections. For several outbreaks in poultry, increased pathogenicity of avian influenza viruses has been linked to the acquisition of multibasic HA cleavage sites, a finding that underscores the significance of the HA cleavage motif for virulence.

Role of PB2 in pathogenicity and host specificity

Recently, the viral replication complex has been recognized as an important contributor to viral pathogenicity, probably by affecting viral growth. The amino acid at position 627 of the PB2 protein was first described as a host range determinant, on the basis of cell culture studies[46]. The respective amino acid change was shown to determine the pathogenicity of H5N1 influenza viruses in mice[47]. Viruses with lysine at this position were pathogenic in mice, whereas those with glutamic acid were non-pathogenic in these animals[47] (Table 2). Notably, almost all human influenza viruses possess lysine at this position, whereas most avian viruses (with the exception of the

Table 2 | Determinants of viral pathogenicity

Protein	Position	Pathogenicity		Swine-origin H1N1 viruses	Function	Reference
		Low	High			
PB2	627	Glu	Lys	Glu	Replicative ability in some mammals, including humans	46, 47
	701	Asp	Asn	Asp	Nuclear import kinetics affecting replicative ability in mice	53, 55
PB1-F2	66	Asn	Ser	Truncated (11 aa)	Induction of apoptosis	74
HA	Cleavage site	Single basic aa	Multiple basic aa	Single basic aa	HA cleavage	45
NS1	92	Asp	Glu	Asp	Unknown (interferon response?)	65
	C terminus	Arg-Ser-Glu-Val, deletion	Glu-Ser-Glu-Val	11 C-terminal aa truncated	Unknown	69, 70

aa, amino acid(s).

'Qinghai Lake' lineage of H5N1 viruses and their descendants) possess glutamic acid at PB2-627. Lysine at position 627 of PB2 is now recognized as a determinant of viral pathogenicity in several mammalian species.

Several studies have addressed the mechanism by which PB2-Lys 627 affects virulence. The amino acid change does not affect tissue tropism in mice but viral replicative ability. This may result from an inhibitory activity in mammalian cells that prevents efficient replication by polymerase complexes possessing PB2-Glu 627 (ref. 48), and/or from inefficient interaction of PB2-Glu 627 with mammalian-type NP protein[49]. Viruses possessing PB2-Lys 627, but not those possessing PB2-Glu 627, grow efficiently in the upper respiratory tract of mammals[50], which may be explained by the fact that PB2-Lys 627 confers efficient replication at 33 °C (the temperature of the upper airway in humans) whereas PB2-Glu 627 does not[51]. In contrast, both variants mediate efficient replication at 37 °C[50,51]. Collectively, these findings suggest that PB2-Lys 627 allows efficient replication not only in the lower, but also in the upper respiratory tract of mammals, a feature that may facilitate transmission. In fact, replacement of PB2-Lys 627 with Glu reduced the transmissibility of human influenza viruses in a guinea-pig model[52].

The amino acid at position 701 of PB2 has also emerged as a determinant of virulence[53,54] (Table 2), a role probably related to its facilitation of binding of PB2 to importin α (a cellular nuclear import factor) in mammalian cells[55]. The recently emerged S-OIVs possess the 'low pathogenic'-type amino acids at positions 627 and 701 (that is, Glu and Asp).

Recently, the PB2 and HA proteins of the Spanish influenza virus were shown to be critical for droplet transmission[13]. The underlying mechanism and the amino acids in PB2 that are critical for this function remain to be determined.

In addition to PB2, other components of the replication complex may also contribute to viral pathogenicity[56]. A recent study also suggested that the replication complex, particularly the PB1 protein, contributes to the virulence of the 1918 pandemic virus in ferrets[12].

Structural data are now becoming available for the viral polymerase complex that may help in the interpretation of mutational analyses; in fact, two studies[57,58] showed that PB2-Lys 627 is part of a basic groove that is disrupted after replacement with Glu.

Role of NS1 in viral pathogenicity

The NS1 protein is an interferon antagonist[59,60] that blocks the activation of transcription factors and IFN-β-stimulated gene products, and binds to double-stranded RNA (dsRNA) to prevent the dsRNA-dependent activation of 2′-5′ oligo(A) synthetase, and the subsequent activation of RNase L, an important player in the innate immune response. Recently obtained structural data are expected to help in the identification of domains that are critical for the biological functions of NS1.

Innate immune responses are stimulated after the recognition of a pathogen by a pathogen-recognition receptor. Several classes of pathogen-recognition receptors have now been described, including retinoic acid inducible gene-I (RIG-I, also known as DDX58) and melanoma differentiation antigen 5 (MDA5, also known as IFIH1),

and Toll-like receptors (TLRs) 3, 7 and 8. Influenza virus infections activate RIG-I signalling, which is counteracted by the viral NS1 protein[61], possibly by forming a complex with RIG-I. Furthermore, influenza virus infection affects TLR7 (refs 62, 63) and TLR4 (ref. 64) signalling, and TLR4-deficient mice do not develop acute lung injury after infection with H5N1 viruses[64]. The direct contribution of NS1 to these signalling events is not known at present.

The NS1 proteins of H5N1 viruses confer resistance to the antiviral effects of interferon and are associated with high levels of pro-inflammatory cytokines[27,28,30,65,66]; the resulting cytokine imbalance probably contributes to the high mortality of H5N1 virus infections in humans. Several amino acids in NS1 have now been shown to affect virulence[65,67,68] (Table 2). The S-OIVs possess the low-pathogenic-type amino acid at these positions. However, available data suggest that these amino acid changes affect virulence in a strain-specific manner, whereas a multibasic HA cleavage sequence and PB2-Lys 627 seem to be universal determinants of viral pathogenicity.

The four carboxy-terminal amino acids of NS1 form a PDZ ligand domain motif that was identified by large-scale sequence analysis[69]. Introduction of the PDZ ligand domains of highly pathogenic H5N1 viruses or the pandemic 1918 virus into an otherwise human virus conferred slightly increased virulence in mice[70] (Table 2). This increase in virulence was not paralleled by increased interferon production. The S-OIVs lack the 11 C-terminal amino acids of NS1, and hence lack the PDZ domain motif. The biological significance of this finding is unknown at present.

Role of PB1-F2 in viral pathogenicity

The PB1 gene of most avian and human influenza A viruses encodes a second protein, PB1-F2, that is expressed from the +1 reading frame[71]. The length of PB1-F2 of swine influenza viruses differs depending on their origin; classical swine viruses possess truncated PB1-F2 proteins of 8–11 amino acids, whereas Eurasian avian-like swine viruses possess full-length PB1-F2 proteins (87–89 amino acids). The S-OIVs encode a truncated PB1-F2 protein of 11 amino acids. PB1-F2 induces apoptosis, probably by interaction with two mitochondrial proteins[71,72], enhances inflammation in mice, and increases the frequency and severity of secondary bacterial infections[15]. It may also affect virulence by interacting with the PB1 protein to retain it in the nucleus for efficient viral replication[73]. A recent study demonstrated that the amino acid at position 66 of PB1-F2 affects the pathogenicity of an H5N1 virus in mice[74] (Table 2). This finding is of great interest because the pandemic 1918 Spanish influenza virus possessed the 'high pathogenic'-type amino acid, and its replacement attenuated this virus[74].

Interspecies transmission

Wild waterfowl are the natural reservoir of influenza A viruses. Besides the continuing transmission of highly pathogenic avian H5N1 influenza viruses to humans, avian influenza A viruses were transmitted to pigs in Europe in 1979, to horses in China, and to seals. Moreover, an avian influenza virus reassorted with a human virus in pigs and transmitted from there to humans.

Human infections with avian or swine influenza viruses have been reported, but they have typically been self-limiting. It is in this context that the recently emerged H1N1 viruses, which are of swine origin and transmit among humans, raise great concern over an imminent pandemic.

Pigs can be naturally and experimentally infected with avian viruses and express both human- and avian-type influenza virus receptors on epithelial cells in trachea[34], supporting the concept of a role as 'mixing vessel' in which human and avian viruses reassort. In North American pig populations, classical swine H1N1 virus dominated for nearly six decades. In 1997–1998, however, H3N2 triple human/avian/swine reassortant viruses emerged that have spread widely within North American pig populations. The emergence of human/avian/swine triple reassortant viruses in pigs further indicates that this species can be infected with human and avian influenza viruses and may provide a platform for reassortment. The triple reassortant viruses that emerged in North American pigs in 1997–1998 are the progenitors of the S-OIVs (Fig. 4).

Prevention and control

For the prevention and control of influenza virus infections, both vaccines and antiviral drugs are available. Nonetheless, the global community is probably not well prepared for a pandemic: antiviral drugs may not be in sufficient supply and the virus may acquire resistance to the available antiviral drugs. On the other hand, the production of a vaccine to a newly emerging strain would take 3–6 months—during which time a virus could spread globally and substantially strain health care systems and the global economy.

Antiviral drugs. Two classes of antiviral drugs—ion channel inhibitors and neuraminidase inhibitors—are at present licensed for use against influenza A viruses.

Adamantanes (that is, amantadine hydrochloride and rimantadine) block the ion channel formed by the M2 protein, which is critical in the release of vRNPs into the cytoplasm. Resistance to adamantanes arises quickly and frequently, and most circulating human H1N1 and H3N2 viruses, some H5N1 viruses, and most European porcine H1N1, H1N2 and H3N2 viruses, are resistant to adamantanes. The S-OIVs are also resistant to ion channel inhibitors[32].

Two neuraminidase inhibitors, oseltamivir and zanamivir, are licensed at present. Neuraminidase inhibitors interfere with the enzymatic activity of the NA protein, which is critical for the efficient release of newly synthesized viruses from infected cells.

In clinical trials, the emergence of resistance to NA inhibitors was rare, and oseltamivir-resistant influenza viruses were attenuated *in vitro* and *in vivo*. Hence, the dissemination of these viruses was not considered an important issue, despite the frequent use of the drug in some countries. However, a study in clinical settings indicated higher (18%) than expected rates of oseltamivir-resistance in children treated with this drug[75]. Recently, the rate of oseltamivir-resistant H1N1 influenza viruses in the United States has increased from 0.7% in the 2006–2007 influenza season to 98.5% in the 2008–2009 influenza season[76]. Similar numbers have been reported for other countries. Equally alarming, oseltamivir-resistant H5N1 viruses have been reported[77,78]. Oseltamivir-resistant human H1N1 viruses may have emerged in immunocompromised patients in which prolonged replication[79–81] may have resulted in the selection of mutations that increase the fitness of oseltamivir-resistant viruses. The S-OIVs are sensitive to neuraminidase inhibitors when tested *in vitro* in enzymatic assays[32]. Recent structural data provide an explanation for oseltamivir-resistance[82], and suggest strategies for the design of improved compounds. In clinical settings, resistance to zanamivir has been reported for an influenza B virus isolated from an immunocompromised child[83].

Several experimental antiviral drugs that target the NA or polymerase proteins are now in different stages of development.

Peramivir, an NA inhibitor that was developed through structure-based design[84], is active in *in vitro* tests against viruses of all nine NA subtypes, including highly pathogenic H5N1 viruses. Phase II clinical trials are now underway to assess the efficacy of intramuscularly administered peramivir against seasonal influenza.

CS-8958, a pro-drug of the new NA inhibitor R-125489 (ref. 85), is a long-acting neuraminidase inhibitor that was found to be effective in phase II clinical trials against seasonal influenza. We also found that CS-8958 has prophylactic and therapeutic efficacy against highly pathogenic H5N1 influenza viruses (M. Kiso, M. Yamashita and Y.K., unpublished data).

T-705 acts as a nucleoside analogue that interferes with the polymerase activity of influenza A, B and C viruses, but also other RNA viruses[86]. It protected mice against infection with highly pathogenic H5N1 viruses (M. Kiso, Y. Furuta and Y.K., unpublished data). Phase II clinical trials for use of T-705 against seasonal influenza viruses have been completed in Japan, and phase III clinical trials are scheduled.

Furthermore, monoclonal antibodies to HA are being developed for the treatment of influenza virus infections. In mice, some antibodies demonstrated prophylactic and therapeutic efficacy against lethal challenge with H5N1 virus[87], suggesting monoclonal antibody treatment as an alternative strategy to treat influenza virus infections.

Vaccines. Seasonal influenza vaccines include human influenza A viruses of the H1N1 and H3N2 subtypes, and an influenza B virus. These vaccines need to be revised every 1–3 years to account for mutations in the HA and NA proteins of circulating viruses (antigen drift).

Inactivated vaccines have been used for many decades. Typically, reassortment is used to generate a seed virus that possesses the HA and NA segments of the circulating virus, and a variable number of segments from A/Puerto Rico/8/34 (H1N1) virus that confer efficient growth in embryonated chicken eggs. The allantoic fluid of embryonated, virus-infected chicken eggs is purified and concentrated by zonal centrifugation or column chromatography, and inactivated with formalin or β-propiolactone. Treatment with detergents or ether and the removal of vRNP complexes leads to split or subunit vaccines that are administered intramuscularly or subcutaneously.

In children and young adults, inactivated influenza vaccines show a 60–80% efficacy rate against laboratory/culture-confirmed influenza illness; this rate is lower in people over the age of 60, that is, the group of people that is most likely to die from influenza virus infections or associated complications. The development of improved influenza virus vaccines is thus clearly warranted.

A live attenuated influenza virus vaccine is now licensed in the United States for healthy individuals aged 2–49. In brief, serial passage of A/Ann Arbor/6/60 (H2N2) virus or B/Ann Arbor/1/66 at low temperatures resulted in viruses that are temperature-sensitive, cold-adapted, and attenuated. These viruses are then reassorted with currently circulating wild-type strains to generate seed viruses that possess the HA and NA genes of the circulating wild-type viruses in the background of the temperature-sensitive, cold-adapted and attenuated phenotypes.

Live attenuated vaccines elicit both humoral and cellular immune responses, and are therefore believed to be superior to inactivated vaccines. In fact, in infants and young children, live attenuated influenza vaccine provides better protection than inactivated vaccine[88]. At present, only egg-based vaccines are licensed for use in the United States. In case of a pandemic, however, eggs may be in short supply. In contrast, cell cultures are highly controllable systems that can be scaled-up for the mass production of vaccines, including those to highly pathogenic H5N1 viruses. The purity and immunogenicity of influenza vaccines produced in Madin-Darby canine kidney or African green monkey kidney cells match those of vaccines produced in embryonated chicken eggs. Cell-culture-based influenza vaccines have been approved for use in humans in Europe.

Recently, particular emphasis has been given to the development of vaccines to highly pathogenic H5N1 viruses. These viruses kill chicken embryos. Propagation of these viruses in eggs therefore results in low yields. To modify these viruses for efficient growth in

embryonated chicken eggs and safe handling by production staff, reverse genetics technologies[5] were used to replace the multibasic HA cleavage site with an avirulent-type HA cleavage site[45], a known virulence factor. Reverse genetics then allowed the generation of a virus possessing the modified HA gene and the NA gene derived from the H5N1 virus and the remaining genes derived from A/Puerto Rico/8/34 virus. Clinical testing of an H5N1 vaccine candidate suggested low immunogenicity[89], prompting the addition of adjuvants to vaccine candidates. In fact, aluminium hydroxide[90,91] or oil-in-water emulsions such as MF59 (refs 92, 93) or AS03 (ref. 94) resulted in considerable antigen-sparing effects. Adjuvanted vaccines also seem to induce broader immune responses, which may be a critical advantage with the emergence of new clades and subclades of H5N1 viruses.

Similarly, the NA and modified HA genes of highly pathogenic avian H5N1 viruses were combined with the remaining genes derived from the live attenuated influenza A virus[95]. To avoid the introduction of H5N1 HA and NA genes into human populations, live attenuated H5N1 vaccines should not be used before a pandemic. However, once H5N1 viruses are widespread in human populations, use of a live attenuated H5N1 vaccine, which is efficacious in non-human primates[96], may be considered to overcome the low immunogenicity of inactivated vaccines.

Several new vaccine approaches are now in various stages of development. A 'universal' vaccine on the basis of the conserved ecto-domain of the M2 protein confers protection against influenza virus infection in animals (reviewed in ref. 97). Various virus-like particles expressing the HA and NA proteins, some in combination with the M1 and M2 proteins, have been tested for their antigenicity and protective efficacy. A recent study demonstrated protection of ferrets against H5N1 virus infection by a virus-like particle expressing the HA, NA and M1 proteins of a heterologous virus[98]. Also, vector approaches have been pursued. In one example, replication-incompetent adeno-viruses expressing an H5 HA protected mice against challenge with homologous and heterologous H5N1 viruses[99,100]. In alternative approaches, Newcastle disease and fowlpox viruses have been explored as vector systems; however, these systems have not been approved for use in humans.

The future

Although much has been learned about influenza viruses, key questions still remain unanswered: for example, what factors determine interspecies transmission, reassortment and human-to-human transmission—factors that have accounted for past pandemics and will be critical in the emergence of new pandemic viruses. The first wave of pandemic Spanish influenza was characterized by relatively low pathogenicity in humans, but the virus presumably mutated into a more virulent form within a few months. Thus, careful monitoring of the S-OIVs during the upcoming winter season in the southern hemisphere is of critical importance to detect more virulent variants, should they arise. From a scientific perspective, the opportunity to watch virus evolution in real time may provide us with invaluable information on the factors that determine pathogenicity, and/or transmissibility. Furthermore, large-scale sequencing efforts, bioinformatics analyses, and the ability to experimentally test predictions with recombinant viruses will eventually provide insight into key features for the emergence of pandemic viruses. In the meantime, the development of improved and new antiviral drugs and vaccines will be critical to control influenza virus outbreaks.

1. Kobasa, D. et al. Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. Nature 445, 319–323 (2007).
2. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. J. Infect. Dis. 198, 962–970 (2008).
3. Taubenberger, J. K., Reid, A. H., Krafft, A. E., Bijwaard, K. E. & Fanning, T. G. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science 275, 1793–1796 (1997).
   This is an important paper that describes the deciphering of the genomic sequence of the 1918 pandemic influenza virus.
4. Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc. Natl Acad. Sci. USA 96, 1651–1656 (1999).
5. Neumann, G. et al. Generation of influenza A viruses entirely from cloned cDNA. Proc. Natl Acad. Sci. USA 96, 9345–9350 (1999).
   This paper describes the artificial generation of influenza viruses, a breakthrough technology that allows the molecular characterization of influenza virus and the generation of influenza virus vaccines.
6. Tumpey, T. M. et al. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310, 77–80 (2005).
   This is a pivotal paper that describes the re-creation of the 1918 pandemic influenza virus.
7. Kash, J. C. et al. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. Nature 443, 578–581 (2006).
8. de Jong, M. D. et al. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nature Med. 12, 1203–1207 (2006).
   This important paper describes high levels of cytokines in humans infected with highly pathogenic avian H5N1 viruses.
9. Kobasa, D. et al. Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus. Nature 431, 703–707 (2004).
10. Tumpey, T. M. et al. Existing antivirals are effective against influenza viruses with genes from the 1918 pandemic virus. Proc. Natl Acad. Sci. USA 99, 13849–13854 (2002).
11. Tumpey, T. M. et al. Pathogenicity and immunogenicity of influenza viruses with genes from the 1918 pandemic virus. Proc. Natl Acad. Sci. USA 101, 3166–3171 (2004).
12. Watanabe, T. et al. Viral RNA polymerase complex promotes optimal growth of 1918 virus in the lower respiratory tract of ferrets. Proc. Natl Acad. Sci. USA 106, 588–592 (2009).
13. Van Hoeven, N. et al. Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. Proc. Natl Acad. Sci. USA 106, 3366–3371 (2009).
14. Geiss, G. K. et al. Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. Proc. Natl Acad. Sci. USA 99, 10736–10741 (2002).
15. McAuley, J. L. et al. Expression of the 1918 influenza A virus PB1–F2 enhances the pathogenesis of viral and secondary bacterial pneumonia. Cell Host Microbe 2, 240–249 (2007).
16. Nakajima, K., Desselberger, U. & Palese, P. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. Nature 274, 334–339 (1978).
   This paper established that the Russian influenza in 1977 was genetically closely related to viruses circulating in humans in the 1950s.
17. Subbarao, K. et al. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science 279, 393–396 (1998).
18. Claas, E. C. et al. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Lancet 351, 472–477 (1998).
19. Smith, G. J. et al. Emergence and predominance of an H5N1 influenza variant in China. Proc. Natl Acad. Sci. USA 103, 16936–16941 (2006).
20. Chen, H. et al. Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. Proc. Natl. Acad. Sci. USA 103, 2845–2850 (2006).
21. Guan, Y. et al. Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. Proc. Natl. Acad. Sci. USA 99, 8950–8955 (2002).
22. Li, K. S. et al. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature 430, 209–213 (2004).
   This paper describes the frequent reassortment events of highly pathogenic avian H5N1 viruses that led to the emergence of the dominant 'genotype Z'.
23. Ducatez, M. F. et al. Avian flu: multiple introductions of H5N1 in Nigeria. Nature 442, 37 (2006).
24. Tran, T. H. et al. Avian influenza A (H5N1) in 10 patients in Vietnam. N. Engl. J. Med. 350, 1179–1188 (2004).
25. The Writing Committee of the World Health Organization (WHO) Consultation on Human Influenza A/H5. Avian influenza A (H5N1) infection in humans. N. Engl. J. Med. 353, 1374–1385 (2005).
26. Chotpitayasunondh, T. et al. Human disease from influenza A (H5N1), Thailand, 2004. Emerg. Infect. Dis. 11, 201–209 (2005).
27. Peiris, J. S. et al. Re-emergence of fatal human influenza A subtype H5N1 disease. Lancet 363, 617–619 (2004).
   This paper describes the re-emergence of human infections with highly pathogenic avian H5N1 viruses in 2003, and also emphasises the high concentrations of cytokines found in infected individuals.
28. To, K. F. et al. Pathology of fatal human infection associated with avian influenza A H5N1 virus. J. Med. Virol. 63, 242–246 (2001).
29. Chan, M. C. et al. Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells. Respir. Res. 6, 135 (2005).

937

30. Cheung, C. Y. *et al.* Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease? *Lancet* **360**, 1831–1837 (2002).

31. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* doi: 10.1126/science.1176062 (in the press).

32. Novel Swine-Origin Influenza A (H1N1) Investigation Team.. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J Med.* doi: 10.1056/NEJMoa0903810 (in the press).
    This highly important paper presents the first summary of epidemiological and virological data on the new swine-origin H1N1 viruses.

33. Rogers, G. N. & Paulson, J. C. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **127**, 361–373 (1983).
    This paper establishes differences between human and avian influenza viruses in receptor-binding specificity.

34. Ito, T. *et al.* Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J. Virol.* **72**, 7367–7373 (1998).

35. Matrosovich, M., Zhou, N., Kawaoka, Y. & Webster, R. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* **73**, 1146–1155 (1999).

36. Shinya, K. *et al.* Avian flu: influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).

37. van Riel, D. *et al.* H5N1 virus attachment to lower respiratory tract. *Science* **312**, 399 (2006).

38. Nicholls, J. M. *et al.* Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract. *Nature Med.* **13**, 147–149 (2007).

39. Stevens, J. *et al.* Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866–1870 (2004).

40. Tumpey, T. M. *et al.* A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* **315**, 655–659 (2007).

41. Gambaryan, A. *et al.* Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology* **344**, 432–438 (2006).

42. Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).

43. Auewarakul, P. *et al.* An avian influenza H5N1 virus that binds to a human-type receptor. *J. Virol.* **81**, 9950–9955 (2007).

44. Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410 (2006).

45. Kawaoka, Y. & Webster, R. G. Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. *Proc. Natl Acad. Sci. USA* **85**, 324–328 (1988).

46. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).

47. Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293**, 1840–1842 (2001).

48. Mehle, A. & Doudna, J. A. An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host Microbe* **4**, 111–122 (2008).

49. Rameix-Welti, M. A., Tomoiu, A., Dos Santos Afonso, E., van der Werf, S. & Naffakh, N. Avian influenza A virus polymerase association with nucleoprotein, but not polymerase assembly, is impaired in human cells during the course of infection. *J. Virol.* **83**, 1320–1331 (2009).

50. Hatta, M. *et al.* Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. *PLoS Pathog.* **3**, e133 (2007).

51. Massin, P., van der Werf, S. & Naffakh, N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J. Virol.* **75**, 5398–5404 (2001).

52. Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog.* **5**, e1000252 (2009).

53. Li, Z. *et al.* Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. *J. Virol.* **79**, 12058–12064 (2005).

54. Gabriel, G. *et al.* Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J. Virol.* **81**, 9601–9604 (2007).

55. Gabriel, G., Herwig, A. & Klenk, H. D. Interaction of polymerase subunit PB2 and NP with importin α1 is a determinant of host range of influenza A virus. *PLoS Pathog.* **4**, e11 (2008).

56. Salomon, R. *et al.* The polymerase complex genes contribute to the high virulence of the human H5N1 influenza virus isolate A/Vietnam/1203/04. *J. Exp. Med.* **203**, 689–697 (2006).

57. Tarendeau, F. *et al.* Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* **4**, e1000136 (2008).

58. Kuzuhara, T. *et al.* Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J. Biol. Chem.* **284**, 6855–6860 (2009).

59. Garcia-Sastre, A. Inhibition of interferon-mediated antiviral responses by influenza A viruses and other negative-strand RNA viruses. *Virology* **279**, 375–384 (2001).

60. Garcia-Sastre, A. *et al.* Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. *Virology* **252**, 324–330 (1998).
    This paper establishes the NS1 protein as an interferon antagonist.

61. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997–1001 (2006).

62. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).

63. Lund, J. M. *et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).

64. Imai, Y. *et al.* Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. *Cell* **133**, 235–249 (2008).

65. Seo, S., Hoffmann, E. & Webster, R. G. Lethal H5N1 influenza viruses escape host anti-viral cytokine responses. *Nature Med.* **8**, 950–954 (2002).

66. Guan, Y. *et al.* H5N1 influenza: a protean pandemic threat. *Proc. Natl Acad. Sci. USA* **101**, 8156–8161 (2004).

67. Jiao, P. *et al.* A single-amino-acid substitution in the NS1 protein changes the pathogenicity of H5N1 avian influenza viruses in mice. *J. Virol.* **82**, 1146–1154 (2008).

68. Li, Z. *et al.* The NS1 gene contributes to the virulence of H5N1 avian influenza viruses. *J. Virol.* **80**, 11115–11123 (2006).

69. Obenauer, J. C. *et al.* Large-scale sequence analysis of avian influenza isolates. *Science* **311**, 1576–1580 (2006).

70. Jackson, D., Hossain, M. J., Hickman, D., Perez, D. R. & Lamb, R. A. A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity. *Proc. Natl Acad. Sci. USA* **105**, 4381–4386 (2008).

71. Chen, W. *et al.* A novel influenza A virus mitochondrial protein that induces cell death. *Nature Med.* **7**, 1306–1312 (2001).

72. Zamarin, D., Garcia-Sastre, A., Xiao, X., Wang, R. & Palese, P. Influenza virus PB1–F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathog.* **1**, e4 (2005).

73. Mazur, I. *et al.* The proapoptotic influenza A virus protein PB1–F2 regulates viral polymerase activity by interaction with the PB1 protein. *Cell. Microbiol.* **10**, 1140–1152 (2008).

74. Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T. & Palese, P. A single mutation in the PB1–F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. *PLoS Pathog.* **3**, e141 (2007).

75. Kiso, M. *et al.* Resistant influenza A viruses in children treated with oseltamivir: descriptive study. *Lancet* **364**, 759–765 (2004).

76. Poland, G. A., Jacobson, R. M. & Ovsyannikova, I. G. Influenza virus resistance to antiviral agents: a plea for rational use. *Clin. Infect. Dis.* **48**, 1254–1256 (2009).

77. Le, Q. M. *et al.* Avian flu: isolation of drug-resistant H5N1 virus. *Nature* **437**, 1108 (2005).

78. de Jong, M. D. *et al.* Oseltamivir resistance during treatment of influenza A (H5N1) infection. *N. Engl. J. Med.* **353**, 2667–2672 (2005).

79. Weinstock, D. M., Gubareva, L. V. & Zuccotti, G. Prolonged shedding of multidrug-resistant influenza A virus in an immunocompromised patient. *N. Engl. J. Med.* **348**, 867–868 (2003).

80. Baz, M., Abed, Y., McDonald, J. & Boivin, G. Characterization of multidrug-resistant influenza A/H3N2 viruses shed during 1 year by an immunocompromised child. *Clin. Infect. Dis.* **43**, 1555–1561 (2006).

81. Ison, M. G., Gubareva, L. V., Atmar, R. L., Treanor, J. & Hayden, F. G. Recovery of drug-resistant influenza virus from immunocompromised patients: a case series. *J. Infect. Dis.* **193**, 760–764 (2006).

82. Collins, P. J. *et al.* Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature* **453**, 1258–1261 (2008).

83. Gubareva, L. V., Matrosovich, M. N., Brenner, M. K., Bethell, R. C. & Webster, R. G. Evidence for zanamivir resistance in an immunocompromised child infected with influenza B virus. *J. Infect. Dis.* **178**, 1257–1262 (1998).

84. Babu, Y. S. *et al.* BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* **43**, 3482–3486 (2000).

85. Yamashita, M. *et al.* CS-8958, a prodrug of the new neuraminidase inhibitor R-125489, shows long-acting anti-influenza virus activity. *Antimicrob. Agents Chemother.* **53**, 186–192 (2009).

86. Furuta, Y. *et al.* In vitro and in vivo activities of anti-influenza virus compound T-705. *Antimicrob. Agents Chemother.* **46**, 977–981 (2002).

87. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nature Struct. Mol. Biol.* **16**, 265–273 (2009).

88. Belshe, R. B. *et al.* Live attenuated versus inactivated influenza vaccine in infants and young children. *N. Engl. J. Med.* **356**, 685–696 (2007).

89. Treanor, J. J., Campbell, J. D., Zangwill, K. M., Rowe, T. & Wolff, M. Safety and immunogenicity of an inactivated subvirion influenza A (H5N1) vaccine. *N. Engl. J. Med.* **354**, 1343–1351 (2006).

90. Bresson, J. L. *et al.* Safety and immunogenicity of an inactivated split-virion influenza A/Vietnam/1194/2004 (H5N1) vaccine: phase I randomised trial. *Lancet* **367**, 1657–1664 (2006).

91. Lin, J. *et al.* Safety and immunogenicity of an inactivated adjuvanted whole-virion influenza A (H5N1) vaccine: a phase I randomised controlled trial. *Lancet* **368**, 991–997 (2006).

92. Bernstein, D. I. *et al.* Effects of adjuvants on the safety and immunogenicity of an avian influenza H5N1 vaccine in adults. *J. Infect. Dis.* **197**, 667–675 (2008).

93. Stephenson, I. *et al.* Antigenically distinct MF59-adjuvanted vaccine to boost immunity to H5N1. *N. Engl. J. Med.* **359**, 1631–1633 (2008).

94. Levie, K. *et al.* An adjuvanted, low-dose, pandemic influenza A (H5N1) vaccine candidate is safe, immunogenic, and induces cross-reactive immune responses in healthy adults. *J. Infect. Dis.* **198**, 642–649 (2008).

95. Suguitan, A. L. Jr *et al.* Live, attenuated influenza A H5N1 candidate vaccines provide broad cross-protection in mice and ferrets. *PLoS Med.* **3**, e360 (2006).

96. Fan, S. *et al.* Immunogenicity and protective efficacy of a live attenuated H5N1 vaccine in nonhuman primates. *PLoS Pathog.* **5**, e1000409 (2009).

97. Schotsaert, M., De, F. M., Fiers, W. & Saelens, X. Universal M2 ectodomain-based influenza A vaccines: preclinical and clinical developments. *Expert Rev. Vaccines* **8**, 499–508 (2009).

98. Mahmood, K. *et al.* H5N1 VLP vaccine induced protection in ferrets against lethal challenge with highly pathogenic H5N1 influenza viruses. *Vaccine* **26**, 5393–5399 (2008).

99. Gao, W. *et al.* Protection of mice and poultry from lethal H5N1 avian influenza virus through adenovirus-based immunization. *J. Virol.* **80**, 1959–1964 (2006).

100. Hoelscher, M. A. *et al.* Development of adenoviral-vector-based pandemic influenza vaccine against antigenically distinct human H5N1 strains in mice. *Lancet* **367**, 475–481 (2006).

Author Contributions G.N. wrote the manuscript. T.N. provided the electron microscopic picture. Y.K. also wrote the manuscript.

ARTICLES

A Jurassic ceratosaur from China helps clarify avian digital homologies

Xing Xu[1], James M. Clark[2], Jinyou Mo[3,4], Jonah Choiniere[2], Catherine A. Forster[2], Gregory M. Erickson[5], David W. E. Hone[1], Corwin Sullivan[1], David A. Eberth[6], Sterling Nesbitt[7], Qi Zhao[1], Rene Hernandez[8], Cheng-kai Jia[9], Feng-lu Han[1,10] & Yu Guo[1,10]

Theropods have traditionally been assumed to have lost manual digits from the lateral side inward, which differs from the bilateral reduction pattern seen in other tetrapod groups. This unusual reduction pattern is clearly present in basal theropods, and has also been inferred in non-avian tetanurans based on identification of their three digits as the medial ones of the hand (I-II-III). This contradicts the many developmental studies indicating II-III-IV identities for the three manual digits of the only extant tetanurans, the birds. Here we report a new basal ceratosaur from the Oxfordian stage of the Jurassic period of China (156–161 million years ago), representing the first known Asian ceratosaur and the only known beaked, herbivorous Jurassic theropod. Most significantly, this taxon possesses a strongly reduced manual digit I, documenting a complex pattern of digital reduction within the Theropoda. Comparisons among theropod hands show that the three manual digits of basal tetanurans are similar in many metacarpal features to digits II-III-IV, but in phalangeal features to digits I-II-III, of more basal theropods. Given II-III-IV identities in avians, the simplest interpretation is that these identities were shared by all tetanurans. The transition to tetanurans involved complex changes in the hand including a shift in digit identities, with ceratosaurs displaying an intermediate condition.

Ceratosaurs are suggested by many recent studies to be closely related to Tetanurae[1,2], within which birds are nested, and they are mainly known from the Cretaceous southern hemisphere[3–6]. Our recent excavations in the Middle–Late Jurassic Shishugou Formation in the Junggar Basin of western China recovered a new ceratosaur, which is one of the earliest known ceratosaurs. This find sheds new light on the morphological evolution in Ceratosauria and in Theropoda as a whole and particularly the digital reduction pattern of theropods.

<div align="center">

Theropoda Marsh, 1881

Ceratosauria Marsh, 1884

Limusaurus inextricabilis gen. et sp. nov.

</div>

Etymology. *Limus*, Latin for mud or mire; *saurus*, Latinization of Greek for lizard; *inextricabilis*, Latin for impossible to extricate. This name is in reference to the specimens' inferred death in a mire.

Holotype. Institute of Vertebrate Paleontology and Paleoanthropology (IVPP) V 15923 is an articulated, nearly complete skeleton (Fig. 1a, b).

Referred material. IVPP V 15924 is a semi-articulated skeleton missing the skull; it is 15% larger than the holotype.

Locality and horizon. Wucaiwan area, Junggar Basin, Xinjiang; Oxfordian upper part of the Shishugou Formation[7].

Diagnosis. Small ceratosaur with the following autapomorphies: short skull (half as long as the femur); skull and mandible toothless; nasal with a lateral shelf; premaxilla with a convex buccal edge; short and wide nasal less than one-third of skull roof length and only twice as long as wide; ventral process of lacrimal strongly inclined anteriorly; slender jugal with rod-like sub-orbital and sub-temporal rami; large external mandibular fenestra about 40% of mandibular length; flange on anterior margin of scapular blade; radius tightly adhering to ulna, and longer than the latter bone; olecranon process absent; metacarpal II much more robust than other metacarpals; metacarpal III with sub-triangular proximal articular surface and non-ginglymoidal distal end; metacarpal I highly reduced and carrying no phalanges; phalanx II-1 with distinct lateral process proximodorsally; pubis with laterally ridged, prominent posterior boot; metatarsus forming a strong transverse arch; robust ventral process at medial margin of proximal end of metatarsal III; metatarsal IV nearly straight, appressed against lateral surface of metatarsal III for nearly its whole length; and pedal digit I small, only 17% as long as metatarsal III.

Morphological description and comparison

Osteological and histological features indicate that both specimens of *Limusaurus inextricabilis* are young adults, probably between the exponential and stationary phases of growth (Fig. 1c; Supplementary Information). It shares some cranial features with both coelophysids and other ceratosaurs and also possess some unique features. *Limusaurus* has a fully developed rhamphotheca. Among non-avian theropods, this condition has been previously reported only in some Cretaceous coelurosaurs[8], so this new find extends the distribution of rhamphothecae within theropods both temporally and phylogenetically.

Postcranially, *Limusaurus* displays a single, fused sternal plate. Unquestionable ossified sternal elements have been previously

[1]Institute of Vertebrate Paleontology and Paleoanthropology, Beijing 100044, China. [2]Department of Biological Sciences, George Washington University, Washington DC 20052, USA. [3]Natural History Museum of Guangxi, Nanning, Guangxi 530012, China. [4]Faculty of Earth Sciences, China University of Geosciences, Wuhan, Hubei 430074, China. [5]Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA. [6]Royal Tyrrell Museum, Drumheller, Alberta T0J 0Y0, Canada. [7]American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA. [8]Instituto de Geologia, Universidad Nacional Autónoma de México, Ciudad Universitaria, Del. Coyocan, México DF 04510, Mexico. [9]Research Institute of Exploration and Development, Xinjiang Oilfield Company, Karamay, Xinjiang 834000, China. [10]Graduate School of Chinese Academy of Sciences, Beijing 100039, China.
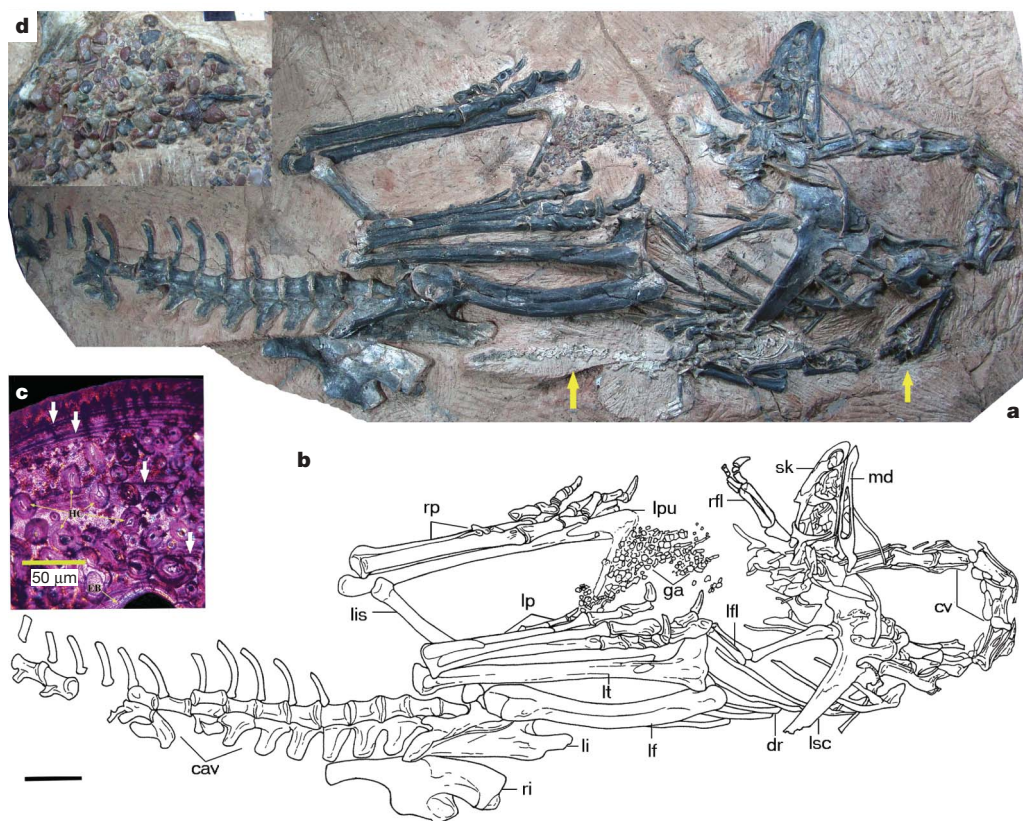
Figure 1 | *Limusaurus inextricabilis* (IVPP V 15923). Photograph (**a**) and line drawing (**b**) of IVPP V 15923. Arrows in **a** point to a nearly complete and fully articulated basal crocodyliform skeleton preserved next to IVPP V 15923 (scale bar, 5 cm). **c**, Histological section from the fibular shaft of *Limusaurus inextricabilis* (IVPP V 15924) under polarized light. Arrows denote growth lines used to age the specimen; HC refers to round haversian canals and EB to layers of endosteal bone. The specimen is inferred to represent a five-year-old individual and to be at a young adult ontogenetic stage, based on a combination of histological features including narrower outermost zones, dense haversian bone, extensive and multiple endosteal bone depositional events and absence of an external fundamental system. **d**, Close up of the gastroliths (scale bar, 2 cm). Abbreviations: cav, caudal vertebrae; cv, cervical vertebrae; dr, dorsal ribs; ga, gastroliths; lf, left femur; lfl, left forelimb; li, left ilium; lis, left ischium; lp, left pes; lpu, left pubis; lsc, left scapulocoracoid; lt, left tibiotarsus; md, mandible; rfl, right forelimb; ri, right ilium; rp, right pes; sk, skull.

reported only in relatively derived coelurosaurs among theropods[9]. A widely arched furcula is present, a feature first reported among ceratosaurs[3]. *Limusaurus* has an abbreviated forelimb as in other ceratosaurs. Metacarpal I is extremely reduced and lacks phalanges, and metacarpal IV is very slender with an unknown number of phalanges (Fig. 2a, b). The elongate hind limbs have a femur + tibiotarsus + metatarsal III/trunk length ratio of 1.80. The tibiotarsus and pes measure about 120% and 130% of the length of the femur, respectively. Similar proportions are seen in derived coelurosaurs[10], and their appearance in *Limusaurus* indicates that strong cursorial capability emerged independently at an early stage of theropod evolution (see Supplementary Information for more morphological description).

Implications for neotheropod evolution

Our phylogenetic analysis places *Limusaurus* in a very basal position within Ceratosauria (Supplementary Information). Some characteristics of *Limusaurus*, such as the hypertrophied scapulocoracoid and highly abbreviated forelimbs with very short hands, were previously considered to diagnose lower-level or even species-level ceratosaurian taxa[3,11,12]. In our analysis, they are optimized as synapomorphies of much more inclusive ceratosaurian groups. Even more significant is the presence in *Limusaurus* of many features also seen in coelophysids and/or tetanurans[9], further reducing the morphological gaps among the three major theropod groups. Features shared with coelophysids are mostly plesiomorphic, but those shared with tetanurans are derived, thus providing further support for a close relationship between Ceratosauria and Tetanurae[3].

Biogeographically, *Limusaurus* is the first definitive ceratosaur known from East Asia[4], to our knowledge, suggesting a cosmopolitan distribution for the group. In combination with other recent discoveries[13], this new ceratosaur makes the Asian dinosaurian fauna less endemic during the Middle–Late Jurassic, suggesting a faunal connection between Asia and other continents during that time period in spite of the presence of the Turgai Sea[14].

Convergent evolution of herbivory

Limusaurus, ornithomimosaurs and shuvosaurid suchians such as *Effigia* are distantly related phylogenetically and also significantly separated temporally[15,16], but they are remarkably similar in many gross skeletal features. They all have a small head with large orbits, toothless upper and lower jaws, a long neck and elongated hind limbs[15]. *Limusaurus* and shuvosaurids also share an extremely large mandibular fenestra and reduced forelimbs. Furthermore, both specimens of *Limusaurus* preserve gastroliths that are similar in quantity, size and shape (Fig. 1d) to those found in some ornithomimid specimens[17]. Another basal ceratosaur, *Elaphrosaurus*, was once actually placed within the Ornithomimosauria[18]. Together with the discovery of *Limusaurus*, this represents an extraordinary case of convergence among three higher archosaurian groups.

Some anatomical features of *Limusaurus* (small toothless head and long neck), and particularly the presence of a gastric mill, indicate a herbivorous diet. Secondary herbivory has previously been documented only in some relatively derived Cretaceous taxa among non-avian theropods[17,19]. As a basal ceratosaur from the Oxfordian, *Limusaurus* represents the earliest and most basal theropod inferred
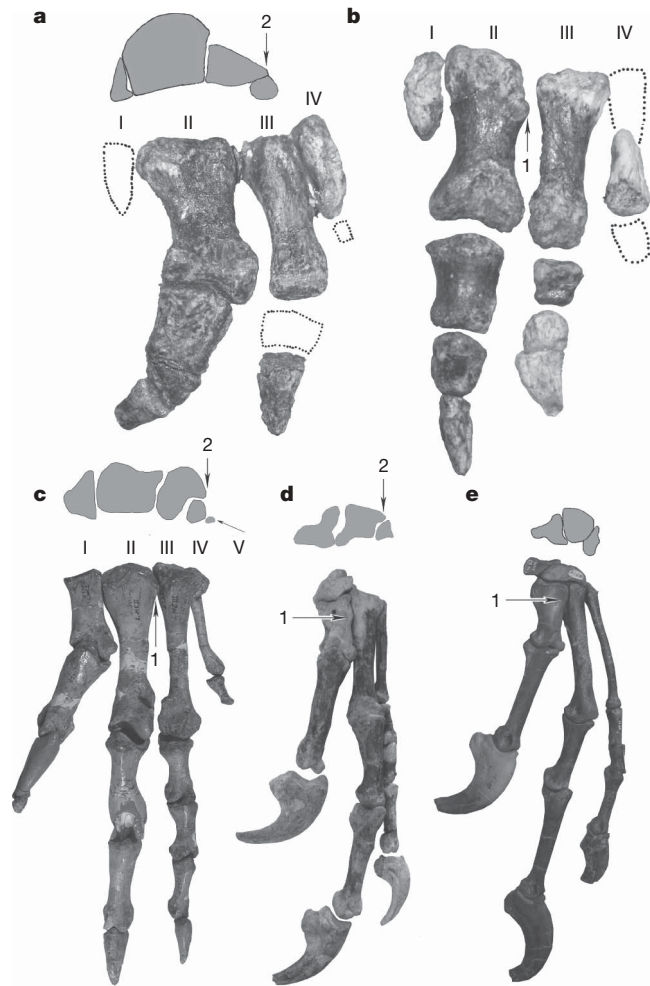
Figure 2 | Theropod manual morphologies as represented by several non-avian theropods. a, b, Ceratosaur *Limusaurus* (IVPP V 15923 and 15924); **c,** basal theropod *Dilophosaurus* (UCMP 37302); **d,** tyrannosauroid *Guanlong* (IVPP V14531); **e,** dromaeosaurid *Deinonychus* (YPM 5206). 1, dorsolateral process; 2, metacarpal IV located ventral to metacarpal III. Note that the three metacarpals of *Guanlong* and *Deinonychus* display many similarities to metacarpals II–IV of *Limusaurus* and *Dilophosaurus*. Interestingly, many metacarpal features, such as the contacts among the three metacarpals and the morphology of the lateral metacarpal, were previously considered to be tetanuran synapomorphies, but in fact can be better interpreted as retained unchanged from the condition in non-tetanuran theropods if the three metacarpals of tetanurans are identified as II-III-IV.

to have been herbivorous, significantly expanding the known trophic diversity of Jurassic theropods.

Manual digit reduction of theropods and avian digital homologies

Theropods have long been assumed to display a pattern of lateral digit reduction (LDR), in which the digits have been progressively reduced from the lateral (that is, ulnar) side of the manus[20–23]. In contrast, bilateral digit reduction (BDR) is characteristic of most other tetrapod groups[23]. However, the reduction of digit I in *Limusaurus* constitutes strong new evidence for BDR in ceratosaurs, particularly because other ceratosaurs also possess a somewhat reduced digit I[24,25].

The occurrence of BDR in Ceratosauria, the sister group of Tetanurae, invites a reconsideration of digital evolution in theropods as a whole, and particularly of the complex issue of tetanuran digital homologies[21–23]. On the basis of morphological evidence from fossil taxa, the three digits retained by tetanurans have traditionally been interpreted as homologues of digits I–III of the primitive theropod

manus, exemplifying LDR[20,21]. However, the discovery of BDR in *Limusaurus* and its close relatives introduces the possibility that this pattern might be more broadly distributed among non-avian theropods and indicates that the three digits of extinct tetanuran theropods could be II–IV, an alternative hypothesis previously little considered in palaeontological literature[26].

Positional relationships have been widely accepted as the main operational criterion for primary homology[27], although cases of positional shifts have been documented[28] or experimentally induced[29]. In the present case, the conservative pentadactyl pattern seen in the embryos of extant birds and crocodilians, and by inference all crown-group archosaurs including theropods, provides a reliable reference system for topologically assessing the primary homologies of tetanuran digits. Recent developmental studies favour the II-III-IV hypothesis by showing that the three digits of the only living tetanurans, the birds, originate developmentally from the middle three of the five digital primordia[23,30–33]. Ontogenetic research on expression patterns of posterior *Hoxd* genes shows that digits that develop from positions II-III-IV in birds acquire a I-II-III identity later in ontogeny[34,35], resolving the apparent conflict between palaeontological and developmental data. Despite the strength of this evidence, developmental data from extant taxa cannot indicate the point at which digital identities shifted during the evolution of the Theropoda, nor reveal the tempo of that shift. The fossil record remains the only source of information on these aspects of the transition. Furthermore, in fossil tetanurans early embryonic stages cannot be observed, so only morphological criteria are available to infer digit identities. Comparing the digits of tetanurans to those of their closest relatives, Ceratosauria and *Dilophosaurus*, is particularly helpful in elucidating digital primary homologies.

The main morphological features cited in support of the traditional I-II-III hypothesis include the topographic relationship of the 'semilunate' carpal to the metacarpus, the short and distally asymmetrical medial metacarpal, and the apparently conserved phalangeal formula of 2-3-4 (refs 20–22, 36). The phalangeal formula is particularly striking because it characterizes digits I-II-III across a wide range of disparate tetrapod groups[21]. However, contrary evidence can be adduced against each of these points. Theropod carpal homologies are complicated by anomalies such as the presence of a relatively small, separate medial distal carpal in non-maniraptoran tetanurans and the absence of a large distal carpal in *Dilophosaurus*[37]. Metacarpal II is distally asymmetric in *Limusaurus*, *Dilophosaurus*[38] and some specimens of the coelophysid *Coelophysis*[39], so the asymmetry of metacarpal I is not a unique feature. Finally, digits I–III do not display a 2-3-4 phalangeal formula in any known ceratosaur, demonstrating that the conservatism of this formula is not absolute.

New information from *Limusaurus* and various other theropods reveals a number of morphological features that support the alternative II-III-IV hypothesis (Fig. 2). In basal theropods the proximal ends of metacarpals I and II are mutually appressed without overlap whereas the dorsolateral corner of the proximal end of metacarpal II forms a flange that slightly overlaps the dorsal surface of metacarpal III. In *Limusaurus* and tetanurans, a similar flange extends distally to form a large, oblique contact between metacarpals II and III in *Limusaurus* and between the medial and middle metacarpals of tetanurans. This indicates that the medial and middle metacarpals of tetanurans correspond to metacarpals II and III. Similarly, the proximal end of metacarpal IV is appressed to the ventrolateral face of metacarpal III in non-tetanuran theropods. The lateral metacarpal of tetanurans contacts the ventrolateral face of the middle metacarpal in the same way, reinforcing the II-III-IV interpretation. The medial metacarpal of basal tetanurans is the most robust in the manus, like metacarpal II in more basal theropods including ceratosaurs, *Dilophosaurus*[38] and coelophysids[3]. The elongate proximal phalanx of the medial digit of tetanurans is similar to phalanx I-1 in some coelophysids but unlike the relatively short phalanx I-1 seen in

ceratosaurs[3], *Dilophosaurus*[38] and *Herrerasaurus*[40]. The middle metacarpal is longer than the others in tetanurans, like metacarpal III in *Limusaurus*, *Dilophosaurus*[38], at least some coelophysids, *Herrerasaurus*[40] and most other archosaurs[26]. The four-fingered ornithischian *Psittacosaurus* displays LDR, and its metacarpal III is also the longest in the manus[41]. A dorsolateral process is present on the proximal end of the middle metacarpal of basal tetanurans, and a similar process characterizes metacarpal III of *Limusaurus*, *Dilophosaurus* and *Herrerasaurus*. Finally, the lateral metacarpal is short, slender and proximally sub-triangular in outline in basal tetanurans, like metacarpal IV in non-tetanuran theropods[3]. It is noteworthy that most of these similarities are more evident between basal tetanurans and their close outgroups, *Dilophosaurus* and the ceratosaurs, than between derived tetanurans and coelophysids.

We conducted a quantitative analysis of digital homologies to test the alternative I-II-III and II-III-IV hypotheses in tetanurans (Supplementary Information). When birds are coded as II-III-IV, coding all Tetanurae as II-III-IV(-V) is a minimum of ten steps shorter than a shift from I-II-III to II-III-IV anywhere within the Tetanurae, and four steps shorter than a shift at the base of the Averostra. Coding all Tetanurae as having II-III-IV is the same in tree length as (characters unordered) or six steps longer than (characters ordered) an alternative scheme in which all Tetanurae, including birds, are interpreted as having I-II-III, a hypothesis that clearly contradicts developmental data from extant birds. We conclude that, if birds possess digits II-III-IV as most developmental studies indicate, the data strongly support the interpretation that all tetanurans have digits II-III-IV(-V), as outlined above. If extant birds are ultimately found to possess digits I-II-III, of course, then no conflict between neontological and traditional palaeontological data exists.

This implies the reduction of digit I before the divergence of the Ceratosauria and the Tetanurae, the appearance of some pollex-like features in digit II and the acquisition of a novel phalangeal formula (X-2-3-4-X) early in tetanuran evolution. Both modifications are partially indicated by the manual morphologies of ceratosaurs and more basal theropods. Also, they are indirectly supported by observations in living animals that a digit will display features normally associated with the neighbouring medial digit if the latter fails to chondrify in early development[21], that phalangeal counts can vary even within species[29,42] and that secondarily cartilaginous elements can regain their ability to ossify[43].

The frameshift hypothesis of digital evolution in theropods holds that each digit, positionally defined, assumed morphological features that primitively characterized the next most medial digit due to homeotic transformations[21]. Recent studies indeed confirm that a homeotic change has affected the development of the avian digits: in extant birds, condensation II receives a Hox signal (absence of posterior *Hoxd* expression) appropriate for digit I[34,35]. The frameshift was, however, not necessarily a sudden, discrete event, and its re-patterning of the digits was not complete, because the metacarpals of basal tetanurans in general retained key features indicating their original identities. The uneven modification of the tetanuran manus may reflect the fact that tetrapod digits develop from proximal to distal[44], each metacarpal appearing before its associated phalanges. The manus may have been repatterned by a late-acting developmental signal that influenced the phalanges to a greater degree than the metacarpals.

In conclusion, both the I-II-III and II-III-IV hypotheses can draw some supporting morphological evidence from the hands of extinct tetanurans, but largely from different manual regions (Fig. 3). If extant tetanurans have retained the middle three digits, as many developmental studies suggest[23,30–33], it is more parsimonious to identify the three digits of extinct tetanurans as digits II–IV. This new evidence from *Limusaurus* and other basal theropods suggests that a gradual homeotic shift in digit identity characterized early stages of theropod evolution, that an intermediate stage of this shift
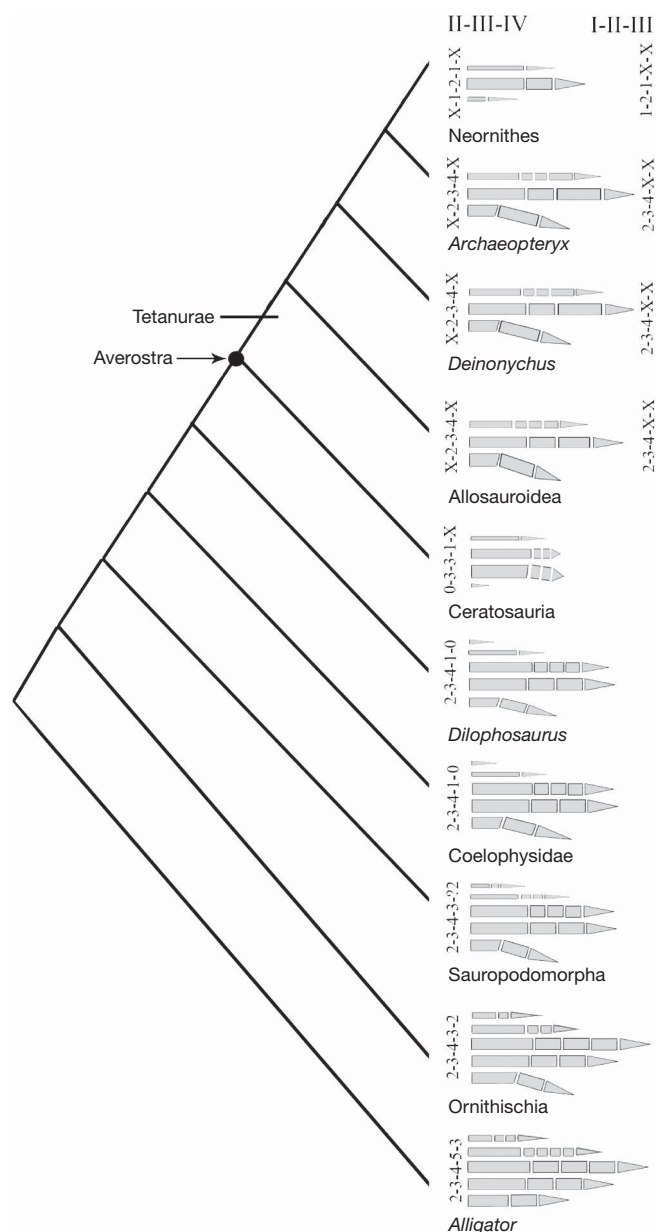
Figure 3 | Manual digital evolution in theropod dinosaurs. Manual digital evolution involves both BDR and LDR in theropod dinosaurs. The shift to BDR in ceratosaurs is coincident with features indicating a reduction in the grasping function of the manus. In ceratosaurs, the manus is small, the manual phalanges are abbreviated and the claws are non-raptorial. This supports the hypothesis that a grasping function constrained the hand to LDR in non-tetanuran theropods[21]. If BDR applies to the more inclusive Averostra, as the II-III-IV hypothesis suggests, early stages of tetanuran evolution must have involved loss of the already highly reduced metacarpal I, reduction in the length of metacarpal II and the reappearance of additional phalanges on metacarpal IV. Both the I-II-III and II-III-IV hypotheses can claim a degree of support from morphological data, but the II-III-IV hypothesis is more parsimonious when developmental data from extant birds are considered.

is preserved in the Ceratosauria and that the shift was complete by the time of the diversification of the earliest tetanurans.

1. Rauhut, O. W. M. *The Interrelationships and Evolution of Basal Theropod Dinosaurs* (Palaeontological Association, 2003).
2. Carrano, M. T., Sampson, S. D. & Forster, C. A. The osteology of *Masiakasaurus knopfleri*, a small abelisauroid (Dinosauria: Theropoda) from the Late Cretaceous of Madagascar. *J. Vertebr. Paleontol.* **22**, 510–534 (2002).

3. Tykoski, R. S. & Rowe, T. in *The Dinosauria* 2nd edn (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 47–70 (Univ. California Press, 2004).

4. Allain, R. *et al.* An abelisauroid (Dinosauria: Theropoda) from the Early Jurassic of the High Atlas mountains, Morocco, and the radiation of ceratosaurs. *J. Vertebr. Paleontol.* **27**, 610–624 (2007).

5. Mateus, O., Walen, A. & Antunes, M. T. The large theropod fauna of the Lourinhã Formation (Portugal) and its similarity to that of the Morrison Formation, with a description of a new species of *Allosaurus. New Mexico Mus. Nat. Hist. Sci. Bull.* **36**, 123–129 (2006).

6. Carrano, M. T. & Sampson, S. D. The phylogeny of Ceratosauria (Dinosauria: Theropoda). *J. Sys. Palaeontol.* **6**, 183–236 (2008).

7. Eberth, D. A. *et al.* Sequence stratigraphy, paleoclimate patterns and vertebrate fossil preservation in Jurassic–Cretaceous strata of the Junggar Basin, Xinjiang Autonomous Region, People's Republic China. *Can. J. Earth Sci.* **38**, 1627–1644 (2001).

8. Chiappe, L. M., Ji, S.-A., Ji, Q. & Norell, M. A. Anatomy and systematics of the Confuciusornithidae (Theropoda: Aves) from the late Mesozoic of northeastern China. *Bull. Am. Mus.* **242**, 1–89 (1999).

9. Holtz, T. R. J., Molnar, R. E. & Currie, P. J. in *The Dinosauria* 2nd edn (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 71–110 (Univ. California Press, 2004).

10. Christiansen, P. & Bonde, N. Limb proportions and avian terrestrial locomotion. *J. Ornithol.* **143**, 356–371 (2002).

11. Bonaparte, J. F. The Gondwanian theropod families Abelisauridae and Noasauridae. *Hist. Biol.* **5**, 1–25 (1991).

12. Sereno, P. C. *et al.* Predatory dinosaurs from the Sahara and Late Cretaceous faunal differentiation. *Science* **272**, 986–991 (1996).

13. Xu, X. *et al.* A basal tyrannosauroid dinosaur from the Late Jurassic of China. *Nature* **439**, 715–718 (2006).

14. Russell, D. A. The role of Central Asia in dinosaurian biogeography. *Can. J. Earth Sci.* **30**, 2002–2012 (1993).

15. Makovicky, P. J., Kobayashi, Y. & Currie, P. J. in *The Dinosauria* 2nd edn (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 137–150 (Univ. California Press, 2004).

16. Nesbitt, S. The anatomy of *Effigia okeeffeae* (Archosauria, Suchia), theropod-like convergence, and the distribution of related taxa. *Bull. Am. Mus. Nat. Hist.* **302**, 1–84 (2007).

17. Kobayashi, Y. *et al.* Herbivorous diet in an ornithomimid dinosaur. *Nature* **402**, 480–481 (1999).

18. Galton, P. M. *Elaphrosaurus*, an ornithomimid dinosaur from the Upper Jurassic of North America and Africa. *Palaeontologische Zeitschrift* **56**, 265–276 (1982).

19. Xu, X., Cheng, Y. N., Wang, X. L., Chang, C. H. & Chang, H. An unusual oviraptorosaurian dinosaur from China. *Nature* **419**, 291–293 (2002).

20. Gauthier, J. in *The Origin of Birds and the Evolution of Flight* (ed. Padian, K.) 1–55 (California Academy of Sciences, 1986).

21. Wagner, G. P. & Gauthier, J. A. 1,2,3 = 2,3,4: a solution to the problem of the homology of the digits in the avian hand. *Proc. Natl Acad. Sci. USA* **96**, 5111–5116 (1999).

22. Shubin, N. H. in *Homology: The Hierarchical Basis of Comparative Biology* (ed. Hall, B. K.) 249–271 (Academic, 1994).

23. Burke, A. C. & Feduccia, A. Developmental patterns and the identification of homologies in the avian hand. *Science* **278**, 666–668 (1997).

24. Burch, S. & Carrano, M. Abelisaurid forelimb evolution: new evidence from *Majungasaurus crenatissimus* (Abelisauridae: Theropoda). *J. Vertebr. Paleontol.* **28** (supplement to 3) 58A (2008).

25. Cortia, R. A., Chiappe, L. M. & Dingus, L. A new close relative of *Carnotaurus sastrei* Bonaparte 1985 (Theropoda: Abelisauridae) from the Late Cretaceous of Patagonia. *J. Vertebr. Paleontol.* **22**, 460–465 (2002).

26. Thulborn, R. A. & Hamley, T. L. The reptilian relationships of *Archaeopteryx. Aust. J. Zool.* **30**, 611–634 (1982).

27. Riedl, R. *Die Ordnung des Lebendigen. Systembedingungen der Evolution* (Parey, 1975).

28. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).

29. Dahn, R. D. & Fallon, J. F. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. *Science* **289**, 438–441 (2000).

30. Larsson, H. C. E. & Wagner, G. P. Pentadactyl ground state of the avian wing. *J. Exp. Zool. B* **294**, 146–151 (2002).

31. Feduccia, A. & Nowicki, Z. The hand of birds revealed by early ostrich embryos. *Naturwissenschaften* **89**, 391–393 (2002).

32. Welten, M. C., Verbeek, F. J., Meijer, A. H. & Richardson, M. K. Gene expression and digit homology in the chicken embryo wing. *Evol. Dev.* **7**, 18–28 (2005).

33. Kundrát, M., Seichert, V., Russell, A. P. & Smetana, K. Pentadactyl pattern of the avian wing autopodium and pyramid reduction hypothesis. *J. Exp. Zool. B* **294**, 152–159 (2002).

34. Vargas, A. O. & Fallon, J. F. Birds have dinosaur wings: the molecular evidence. *J. Exp. Zool. B* **304**, 86–90 (2005).

35. Vargas, A. O., Kohlsdorf, T., Fallon, J. F., Brooks, J. V. & Wagner, G. P. The evolution of *HoxD-11* expression in the bird wing: insights from *Alligator mississippiensis. PLoS ONE* **3**, e3325 (2008).

36. Chatterjee, S. Counting the fingers of birds and dinosaurs. *Science* **280**, 355 (1998).

37. Chure, D. J. in *New Perspectives on the Origin and Early Evolution of Birds* (eds Gauthier, J. A. & Gall, L. F.) 122–130 (Yale Univ. Press, 2001).

38. Welles, S. P. *Dilophosaurus wetherilli* (Dinosauria, Theropoda), osteology and comparisons. *Palaeontgr. Abt. A* **185**, 85–180 (1984).

39. Galton, P. M. Manus movements of the coelurosaurian dinosaur *Syntarsus* and opposability of the theropod hallux. *Arnoldia (Rhodesia)* **5**, 1–8 (1971).

40. Sereno, P. C. The pectoral girdle and forelimb of the basal theropod *Herrerasaurus ischigualastensis. J. Vertebr. Paleontol.* **13**, 425–450 (1993).

41. You, H. L. & Dodson, P. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 478–493 (Univ. California Press, 2004).

42. Drossopoulou, G. *et al.* A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* **127**, 1337–1348 (2000).

43. Baksh, D., Boland, G. M. & Tuan, R. S. Cross-talk between Wnt signaling pathways in human mesenchymal stem cells leads to functional antagonism during osteogenic differentiation. *J. Cell. Biochem.* **101**, 1109–1124 (2007).

44. Shapiro, M. D. Developmental morphology of limb reduction in *Hemiergis* (Squamata: Scincidae): chondrogenesis, osteogenesis, and heterochrony. *J. Morphol.* **254**, 211–231 (2002).

Author Contributions X.X. and J.M.C. designed the project. X.X., J.M.C., J.C., G.M.E., S. N. and J.-Y.M. performed the research. X.X., J.M.C., G.M.E., J.C., C.S. and D.W.E.H. wrote the manuscript. X.X., J.M.C., J.-Y.M., J.C., C.A.F., D.A.E., Q.Z., R. H., C.-K. J., F.-L.H. and Y.G. excavated the specimens.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.X. (xingxu@vip.sina.com).

ARTICLES

A newly discovered protein export machine in malaria parasites

Tania F. de Koning-Ward[1,2], Paul R. Gilson[1,3], Justin A. Boddey[1], Melanie Rug[1], Brian J. Smith[1], Anthony T. Papenfuss[1], Paul R. Sanders[1], Rachel J. Lundie[1], Alexander G. Maier[1], Alan F. Cowman[1] & Brendan S. Crabb[1,3]

Several hundred malaria parasite proteins are exported beyond an encasing vacuole and into the cytosol of the host erythrocyte, a process that is central to the virulence and viability of the causative *Plasmodium* species. The trafficking machinery responsible for this export is unknown. Here we identify in *Plasmodium falciparum* a translocon of exported proteins (PTEX), which is located in the vacuole membrane. The PTEX complex is ATP-powered, and comprises heat shock protein 101 (HSP101; a ClpA/B-like ATPase from the AAA+ superfamily, of a type commonly associated with protein translocons), a novel protein termed PTEX150 and a known parasite protein, exported protein 2 (EXP2). EXP2 is the potential channel, as it is the membrane-associated component of the core PTEX complex. Two other proteins, a new protein PTEX88 and thioredoxin 2 (TRX2), were also identified as PTEX components. As a common portal for numerous crucial processes, this translocon offers a new avenue for therapeutic intervention.

Malaria remains one of the world's most important infectious diseases and is responsible for enormous mortality and morbidity[1]. The most lethal form of human malaria is caused by the protozoan parasite *P. falciparum*. Central to the capacity of this organism to grow inside red blood cells and to thrive inside the blood-stream is its ability to export ~5% of its encoded genome (200–300 proteins) into the cytosol of its host cell[2–5]. Although the function of most of these exported proteins is unknown, those that have been investigated appear to have virulence-associated roles, such as promoting infected cell adhesion and/or rigidity[6–9]. Apart from these functions, a recent gene knockout screen of 85 exported proteins implicated an essential blood-stage survival role for ~25% of exported proteins[9]. Moreover, parasite proteins are also exported beyond a vacuolar membrane during the liver stage of infection[10].

To gain access to the host cell cytosol, exported parasite proteins must cross two membranes, the parasite plasma membrane and the parasitophorous vacuole membrane (PVM) that envelops the parasite. Despite its importance, the mechanism of protein export is not known although export initially requires proteins to enter into the secretory pathway[11–13]. This is mediated by a recessed amino-terminal hydrophobic endoplasmic reticulum (ER) signal sequence, which is sufficient to transport proteins beyond the plasma membrane to, but not across, the PVM[12]. For most exported proteins, trafficking across the PVM requires an additional sequence element, known as a *Plasmodium* export element (PEXEL) or a vacuolar transport signal (VTS), which is located approximately 25–30 amino acids downstream of the ER signal sequence[2,3]. The discovery of this element was a major advance, as it allowed predictions of the exported proteome of *P. falciparum* and other *Plasmodium* species[2–5], and it also suggested the presence of a central portal through which most or all exported proteins must pass[2,3].

Identification of two candidate translocon proteins

Export of PEXEL-containing proteins most probably requires a proteinaceous translocon within the PVM[2,3,14], but homology searches for relatives of known members of translocon systems have failed to predict its identity. As an alternative, we combined two approaches to identify candidates: proteomic analysis of relevant parasite membranes and the establishment of predictive criteria. With respect to the first approach, we found that lipid-raft-like detergent resistant membranes from ring-stage parasites were strongly enriched in proteins that are known (or suspected) to localize to the PVM (Supplementary Table 1). Hence, it seemed likely that translocon components would be represented amongst these proteins.

On the basis of existing knowledge, we established five specific criteria to systematically predict putative PEXEL-protein translocon components (Supplementary Table 2). Briefly, we postulated that such proteins should: (1) be restricted to the *Plasmodium* genus only because PEXEL motifs appear to be absent from even closely related Apicomplexan parasites, (2) incorporate a power source, which previous evidence suggests is likely to be an ATPase[15], (3) have dual apical merozoite and ring-stage PVM localization and also be expressed in the liver stage, (4) be essential for blood-stage growth and (5) specifically bind to their exported protein cargo.

The *P. falciparum* ring-stage detergent resistant membrane proteome was analysed for potential translocon components. Of interest was HSP101 (PF11_0175; Supplementary Fig. 1), which belongs to the dual AAA+ ATPase domain containing HSP100/ClpA/B chaperone family. HSP100 proteins are core components of numerous protein translocon systems, such as type VI secretion in Gram-negative bacteria[16] and the translocon at the inner membrane of chloroplasts[17]. Importantly, *Plasmodium* HSP101 encodes an N-terminal ER signal sequence for export into the parasitophorous vacuole and orthologues were only found within the *Plasmodium* genus, as expected for a role in PEXEL-protein trafficking (Supplementary Fig. 1). Members of the HSP100 family form hexameric ring-shaped complexes that generally translocate proteins through a central pore of the ring in an ATP-dependent manner. Structural modelling of *P. falciparum* HSP101 predicts that it forms a typical ring-shaped hexamer (Supplementary Fig. 1).

[1]The Walter & Eliza Hall Institute of Medical Research, Melbourne 3052, Australia. [2]Deakin University, Waurn Ponds, Victoria 3217, Australia. [3]Macfarlane Burnet Institute for Medical Research & Public Health, Melbourne 3004, Australia.

P. falciparum HSP101 has an identical late schizont/early ring transcriptional profile to another protein that was also present in both the ring- (Supplementary Table 1) and schizont-stage proteomes of *P. falciparum* (Supplementary Fig. 2). This hypothetical protein, which we have termed PTEX150 (PF14_0344), also has a putative ER signal sequence and is found throughout the *Plasmodium* genus but not in other genera (Supplementary Fig. 2). Moreover, data from our previous proteomic study identifying large molecular mass complexes in *Plasmodium* membranes demonstrated that HSP101 and PTEX150 were present in two distinct high molecular mass (~440 and >650 kDa) bands[18] (Supplementary Fig. 2), suggesting they form a large complex.

A series of reagents were generated, including PTEX150- and HSP101-specific antibodies and two transgenic *P. falciparum* (3D7) parasite lines termed 3D7-150HA and 3D7-101HA, where the respective endogenous genes were modified at their carboxy termini to include a triple haemagglutinin (HA) epitope tag (Fig. 1a, Supplementary Fig. 3). Antibodies raised against *P. falciparum* PTEX150 recognized a single protein species of 150 kDa in parasite extracts (predicted molecular mass is 112 kDa), while a slightly larger species of 155 kDa was observed in 3D7-150HA due to the addition of the tag (Fig. 1a). The PTEX150 antibody was further verified by its specific reactivity with the 155-kDa HA-tagged species (Fig. 1a). Similarly, antibodies raised against HSP101 recognized a single protein species of the expected molecular masses in both 3D7 and 3D7-101HA parasites (Fig. 1b). The temporal expression patterns of HSP101 and PEX150 showed that both proteins were strongly expressed in late schizogony and remained at similar levels through the 48-hour blood-stage cycle (Fig. 1c). As transcription of these proteins peaks in late schizogony/early ring stages (www.PlasmoDB.org), this indicates that these proteins are stable, with little protein turnover.

Immunofluorescence analysis (IFA) showed that PTEX150 and HSP101 co-localize and are found in discrete foci in the membranes surrounding the ring-stage parasite (Fig. 1d). This membrane is the PVM, as fluorescence surrounds the parasite membrane marker MSP1(19). Also, it is predicted that the PEXEL protein trafficking machinery would already have been made in the merozoite and injected into the newly forming vacuole during erythrocyte invasion, so that it can immediately traffic similarly injected proteins, such as the ring-infected surface antigen (RESA). Consistent with this, both PTEX150 and HSP101 are strongly expressed late in schizogony (Fig. 1c) and reside at the apical end of free merozoites, presumably in secretory organelles (Fig. 1d). Hence, PTEX150 and HSP101 fulfil the first three criteria expected of a PEXEL-protein translocon. With respect to the fourth criterion, we have been unsuccessful in repeated

Figure 1 | HSP101 and PTEX150 co-localize and have dual apical merozoite and PVM localization. a–c, Western blot analysis of parasite proteins extracted from parental 3D7 and transgenic 3D7-150HA parasites (**a**), 3D7 and transgenic 3D7-101HA parasites (**b**) and 3D7-150HA parasites harvested at mixed schizont/ring (S/R), ring (R), early trophozoite (ET), late trophozoite (LT) and early schizont (ES) stages or from mature magnet purified schizont (S) stages, using the indicated antibodies (**c**). AMA1 represents a marker for a protein expressed only at the late stages of parasite development. Molecular mass standards (kDa) are shown on the left. **d,** Double labelling IFA on fixed 3D7-150HA and 3D7-101HA ring- and merozoite-stage parasites using the antibodies as indicated. The arrows indicate the apical end of the merozoite. BF, bright-field; DAPI, nucleic acid stain 4′,6-diamidino-2-phenylindole. The original magnification for all images was ×1,000.

Figure 2 | Isolation of a five-member PTEX complex. a, b, Left, Coomassie-stained SDS–PAGE gels of material eluted from immunoprecipitations performed using HA antibodies on parasite lysates from 3D7 and 3D7-150HA (**a**) or 3D7 and 3D7-101HA (**b**). Right, the total number of peptides present in all excised bands for those parasite-specific proteins that were the top peptide hit in the visible bands unique to pull-downs using transgenic parasites (labelled (1)–(5)), and the peptide abundance of Pfcdc48 and all other parasite and host cell proteins identified. **c,** Immunoblot analysis of immunoprecipitations performed on 3D7 parental and transgenic HA-tagged parasite lysates using antibody combinations for immunoprecipitation (IP Ab) and western blot (WB) as indicated. **d,** Double-labelling IFA on fixed 3D7-150HA ring-stage parasites using a mixture of antibodies specific for PTEX150 and EXP2. Original magnification, ×1,000. **e,** Schematic of the five members of the PTEX complex, showing that each possesses an ER signal sequence and other distinguishing features, such as regions that are highly conserved across the genus.

946

attempts to generate gene knockouts of *P. falciparum* PTEX150 and its rodent malaria orthologue *Plasmodium berghei* PTEX150 (data not shown), suggesting that this protein is essential to blood-stage development.

Identification of a putative translocon complex, PTEX

To investigate whether PTEX150 and HSP101 associate and to identify other potential binding partners, pull-down experiments were performed using HA antibodies on 3D7-150HA and 3D7-101HA parasites. To control for nonspecific interactions, precipitations were also performed on the 3D7 parent (Fig. 2a, b). SDS–PAGE (SDS–polyacrylamide gel electrophoresis) bands visually unique to pull-downs using transgenic parasites by Coomassie staining were excised, as were the corresponding gel regions of the negative control, and analysed by LC–MS/MS (liquid chromatography coupled with tandem mass spectrometry) based sequencing (see bands 1–5 in Fig. 2a, b, and Supplementary Table 3). These data indicated that HSP101 and PTEX150 co-precipitated in a specific and reciprocal manner with 150HA (Fig. 2a) and 101HA (Fig. 2b), respectively. Combined with western-blotted extracts of HA-immunoprecipitations (Fig. 2c), these data confirmed that they form a stable complex. Three additional proteins, identified as the top parasite-specific peptide hit in bands 3–5, were also specifically affinity purified with both 150HA and 101HA species (Fig. 2a, b, and Supplementary Table 3). These proteins include one one known PVM protein EXP2 (PF14_0678)[19], a novel protein termed PTEX88 (PF11_0067) and a thioredoxin-like, TRX2 protein (MAL13P1.225)[20]. The interactions of these proteins were specific, as either none or very few peptides were recovered from pull-downs performed in parallel on parental parasites (Supplementary Table 3). To determine whether other less abundant proteins that co-precipitate with PTEX150 and HSP101 were present, the remaining bands were excised from the gel and analysed by mass spectrometry (Supplementary Table 4). No other proteins specific to both 150HA and 101HA pull-downs were found.

EXP2 has been previously characterized in *P. falciparum* as a 35-kDa ring-stage protein that has been demonstrated to be associated with the PVM[19,21]. By immunoprecipitation, we confirmed that EXP2 interacts with the PTEX150 complex (Fig. 2c). Furthermore, dual labelling IFA experiments demonstrated that EXP2 co-localizes with PTEX150 and HSP101 in large foci in the ring-stage PVM (Fig. 2d). Similar experiments were not performed for PTEX88 or TRX2 as specific reagents are not available. Although we are cautious about assigning these members to the PTEX complex, we emphasize that both proteins have ER signal sequences, are co-regulated at a messenger RNA level with other PTEX members, and are unique to *Plasmodium* (Fig. 2e, Supplementary Fig. 4). In addition, PTEX88 and TRX2 were specifically detected in both PTEX150 and HSP101 affinity purification experiments. Hence both are likely components of the PTEX complex. This is somewhat surprising in the case of TRX2, as a GFP-tagged version of this protein has been previously shown to localize to the mitochondrion[20]. However, the presence of an ER signal sequence and the absence of a mitochondrial-targeting motif in TRX2 are consistent with a location outside the mitochondrion. Indeed, while the TRX2–GFP fusion protein described by Boucher and colleagues localizes to the mitochondrion in the majority of parasites, in others it shows a pattern surrounding the parasite consistent with a parasitophorous vacuole localization (S. Muller, personal communication). It is possible that the heterologous heat shock promoter used for this transgene experiment contributed to this dual localization. Hence, although specific antibodies are required to firmly resolve the localization of TRX2, existing data are consistent with it comprising a component of the PTEX machinery.

In summary, we have identified a core macromolecular complex comprising HSP101, PTEX150 and EXP2 with two additional, potentially accessory, proteins, PTEX88 and TRX2 (Fig. 2e). Each

of these is found throughout the *Plasmodium* genus and obvious orthologues are not present in any other organism. We note, however, that reciprocal BLAST analysis suggests that EXP2 is related to the dense granule family of secreted proteins known as DG32 in other Apicomplexan parasites[22], although the relatively low level of identity and absence of conserved cysteine residues means it is unclear if DG32 proteins represent orthologues of EXP2. Hence, the PTEX complex is a novel PVM-associated, ATPase machine that satisfies the first four essential criteria for a PEXEL-protein translocon.

PTEX specifically interacts with exported proteins

We investigated the capacity of PTEX proteins to interact with their anticipated PEXEL-protein cargo. Proteomic analysis of the affinity purifications performed on 3D7-150HA parasite lysates revealed that while 53% of the peptides identified were PTEX components, another 8% were from exported proteins, presumably the translocon's
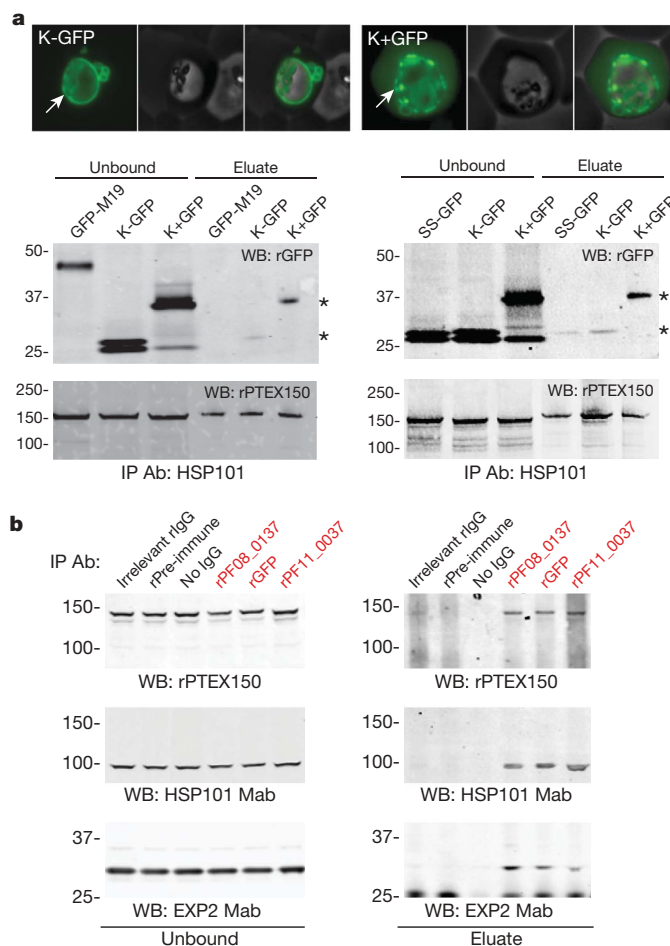
Figure 3 | The PTEX complex interacts with PEXEL proteins. a, Top panels, GFP fluorescence of non-exported (K-GFP, left) versus exported (K+GFP, right) reporters in live transgenic parasites. Arrows highlight the differences in fluorescence at the parasitophorous vacuole, with the exported protein displaying a 'necklace of beads' appearance. Bottom panels, western-blot analysis of immunoprecipitations with HSP101 antibodies using transgenic parasites, which express various GFP reporter proteins (Supplementary Fig. 5). Despite equivalent levels of interaction of PTEX150 with HSP101 in all transgenic parasite lysates (bottom panels), greater amounts of the exported K+GFP protein (asterisks) are pulled down in comparison to any of the non-exported reporter proteins. Original magnification, ×1,000.
b, Immunoprecipitations in the K+GFP transgenic line using antibodies against three exported proteins (shown in red). Western blots of proteins present in unbound (left panel) or eluate (right panel) fractions were probed with antibodies against PTEX components. No antibody, pre-immune and irrelevant (anti-AMA1) IgG were included as negative controls. r, polyclonal antibody raised in rabbits; Mab, monoclonal antibody raised in mice.

cargo (Supplementary Table 4). No peptides of exported proteins were recovered from control 3D7 parasite lysates analysed in parallel, suggesting a specific interaction between PTEX components and exported proteins. The PEXEL protein peptides recovered were from just two, presumably abundant, exported proteins, RESA (PFA0110w) and a previously uncharacterized protein PF08_0137 (Supplementary Fig. 5). To confirm the specificity of this interaction, we performed the reverse experiment using RESA and PF08_0137 antibodies to pull-down cross-linked parasite protein extracts. Under these conditions, both antibodies specifically pulled-down PTEX components HSP101 and PTEX150 to a level 4–6-fold above background levels of binding seen in the bead-only negative control (Supplementary Fig. 5). To further address whether exported proteins specifically interact with the PTEX complex, we tested whether HSP101 antibodies could precipitate different GFP-tagged reporter proteins. In two separate experiments, we show that the exported GFP reporter protein K+GFP[12] showed superior binding to HSP101 (between 4- and 6-fold in different experiments) than the non-exported GFP-reporters (Fig. 3a). These negative controls include a reporter associated with the parasite membrane, GFP-M19 (ref. 23), and two soluble reporters that localize to the parasitophorous vacuole, SS-GFP[24] and K-GFP[12] (Supplementary Fig. 5). Despite possessing a PEXEL motif that is recognized and cleaved in the ER[25], the K-GFP fusion protein is not exported owing to the lack of a spacer region between the PEXEL element and the GFP reporter[26]. To further test the specificity of this and other PEXEL binding protein interactions with the PTEX complex, we

performed the immunoprecipitations in the K+GFP expressing parasite line with antibodies against GFP, PF08_0137 and an additional PEXEL protein PF11_0037 (ref. 9). In these experiments, K+GFP, PF11_0037 and PF08_0137 exported proteins pulled down the PTEX components HSP101, PTEX150 and EXP2 far more strongly (5–200 fold) than did the negative control pre-immune IgG (Fig. 3b). In summary, we have shown that four different PEXEL-motif containing exported proteins demonstrate specific binding to the PTEX machinery.

A model for PTEX function

So how might this apparatus function? Protease sensitivity under differential membrane permeabilization conditions suggests that HSP101 and PTEX150 reside entirely on the vacuolar face of the PVM (Supplementary Fig. 6). Indeed, it is already known that EXP2 predominately resides in this location[21]. Hence, we propose that the PTEX complex is injected directly from apical merozoite organelles into the vacuole, where it associates with the PVM to form large macromolecular foci. It has been shown that PEXEL-containing proteins are recognized and cleaved after the leucine residue in the RxLxE/Q/D PEXEL motif in the parasite's ER[25,27]. This processing step is probably to free the cargo from the ER membrane[27]. The xE/Q/D remaining after cleavage is essential for mature cargo to reach the host cell. This may involve either recognition of this sequence in the PV by the translocon and/or the interaction of xE/Q/D with proteins in the ER or Golgi that sort cargo into vesicles that traffic proteins to the specific regions of the PV that contain the PTEX translocon[27]. Once in the vacuole, proteins destined for export are known to be unfolded presumably before being fed through a membrane pore[28]. HSP101 is most likely to be responsible for unfolding and feeding denatured proteins through its central channel, but the identity of the transmembrane pore itself is not obvious. Although it does not possess a predicted transmembrane domain, evidence that EXP2 is membrane-associated has been published[19,21]. Here we show that EXP2 is the most resistant of the three core PTEX components to extraction with carbonate, consistent with it being the direct membrane-associated component of the PTEX complex (Fig. 4a). Notably, using the FUGUE structural homology software[29], EXP2 is predicted to fold similarly to the pore forming toxin *Escherichia coli* haemolysin E (HLyE; Supplementary Fig. 7). Although it is the 'top hit' in this analysis, the predictive score is in the 'moderate' range of confidence, and other structural homology searches did not yield significant hits. However, adding weight to the prospect that EXP2 folds similarly to HLyE is the prediction in the structural homology model of EXP2 (built from the HLyE structure) of an intramolecular disulphide bond; this C24-C113 bond in EXP2 is predicted to be in a very similar location to the conserved cysteine pair at the base of HLyE[30] (Supplementary Fig. 7). Indeed, EXP2 must fold in a manner that allows intramolecular disulphide bond(s), as EXP2 migration by SDS–PAGE is sensitive to reduction[21]. Taken together, EXP2 is the most likely of the identified components to constitute the membrane pore, although we note that direct evidence for this is lacking at present; indeed, it remains possible that other, perhaps yet to be identified, PTEX components perform this role. (See Fig. 4b for a model of PEXEL-protein translocation.)

The function of the other PTEX components is unclear, although scaffolding and cargo-binding functions are likely for some members. TRX2 has a number of potential roles, including aiding the unfolding of PEXEL-proteins before passage through the translocon or regulating PTEX function. As for the red blood cell cytosolic side of the membrane, it is notable that we detected a six-fold enrichment of human HSP70 family protein, HSC70, in pull-downs from the HA-tagged parasites compared to parental parasites (Supplementary Table 4). Host cell HSP70 may be recruited to the translocon, perhaps in conjunction with exported parasite proteins (such as PF08_0137, which we also suggest binds to the PTEX complex; Fig. 3 and Supplementary Fig. 5), to assist in the passage and/or refolding of
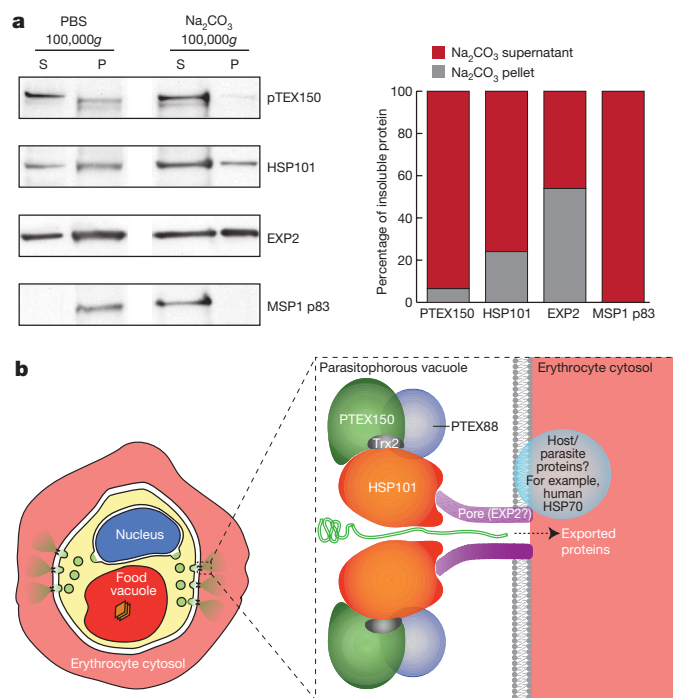
Figure 4 | Model for PTEX function. a, Differential solubilization of schizont-stage parasites to analyse membrane association. Parasites were first solubilized in PBS into supernatant (S) and pellet (P) fractions and analysed by western blotting. The PBS insoluble pellet was further solubilized in Na_2CO_3 and similarly analysed. The p83 fragment of MSP1 is included as a control for solubilization of a peripherally associated membrane protein. **b,** Proposed model for PEXEL-protein export. Once deposited into the vacuolar space, proteins destined for export are recognized by some member(s) of the PTEX complex and deposited into the N-terminal domain of HSP101 where they are unfolded. These proteins are fed through the central channel of HSP101 and ultimately through a membrane-associated channel, predicted here to be EXP2. On the cytosolic face, host (such as human HSP70) and/or exported parasite proteins may assist in energizing translocation and/or in refolding proteins.

exported proteins. Consistent with such a role, erythrocyte HSP70 transforms from a soluble to a membrane-bound state in parasitized cells[31].

Although many mechanistic questions surrounding our model for *Plasmodium* protein export remain to be addressed, this study reveals the identity of a novel protein trafficking apparatus that serves as a common nexus for proteins involved in virulence and survival, and hence provides a 'Achilles heel' for anti-malaria drug developers. The evolutionary origin of this machinery remains unknown, as it is absent from even closely related parasites and is distinct from other characterized protein translocon machineries, such as the secretion systems in bacterial cell membranes or translocons in eukaryotic organelles. However, the conserved use of hexameric ring-forming AAA+ ATPases in many of these systems, including the PTEX apparatus we describe here, hints at an ancient origin of this key biological process.

METHODS SUMMARY

Transgenic lines 3D7-150HA and D37-101HA were generated by transfecting *P. falciparum* strain 3D7 with plasmids pPTEX150-HA/Str3' or pHSP101-HA/Str3', respectively. Parasites containing integrated forms of the plasmid were obtained by repeated drug cycling in the presence and absence of WR99210. Antibodies were also raised against recombinant hexa-His fusion proteins derived from either PTEX150 (amino acids 181–236) or HSP101 (amino acids 68–170). These transgenic lines and antibodies, along with other antibodies described in this study, were used in immunoprecipitation experiments as well as western blot and immunofluoresence analysis. Immunoprecipitations were performed on mixed ring and trophozoite-stage infected erythrocytes in which the erythrocyte had been permeabilized with tetanolysin or first cross-linked with 2 mM DSP followed by saponin lysis. Proteins present in parasite lysates (generated by incubation of the parasites on ice in 1% (v/v) Triton X-100 or RIPA buffer) that were pulled down with anti-HA antibodies or other antibodies described in this study were identified either by LC–MS/MS based sequencing or by western blot analysis. For IFA, parasites were fixed with ice-cold 90% acetone/10% methanol before incubation with various primary and secondary antibody combinations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 30 April 2009.

1. Sachs, J. & Malaney, P. The economic and social burden of malaria. *Nature* **415**, 680–685 (2002).
2. Marti, M. *et al.* Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
3. Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
4. Sargeant, T. J. *et al.* Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* **7**, R12 (2006).
5. van Ooij, C. *et al.* The malaria secretome: from algorithms to essential function in blood stage infection. *PLoS Pathog.* **4**, e1000084 (2008).
6. Crabb, B. S. *et al.* Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* **89**, 287–296 (1997).
7. Waterkeyn, J. G. *et al.* Targeted mutagenesis of *Plasmodium falciparum* erythrocyte membrane protein 3 (PfEMP3) disrupts cytoadherence of malaria-infected red blood cells. *EMBO J.* **19**, 2813–2823 (2000).
8. Glenister, F. K. *et al.* Contribution of parasite proteins to altered mechanical properties of malaria-infected red blood cells. *Blood* **99**, 1060–1063 (2002).
9. Maier, A. G. *et al.* Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell* **134**, 48–61 (2008).
10. Singh, A. P. *et al. Plasmodium* circumsporozoite protein promotes the development of the liver stages of the parasite. *Cell* **131**, 492–504 (2007).
11. Benting, J., Mattei, D. & Lingelbach, K. Brefeldin A inhibits transport of the glycophorin-binding protein from *Plasmodium falciparum* into the host erythrocyte. *Biochem. J.* **300**, 821–826 (1994).
12. Wickham, M. E. *et al.* Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J.* **20**, 5636–5649 (2001).
13. Lopez-Estraño, C., Bhattacharjee, S., Harrison, T. & Haldar, K. Cooperative domains define a unique host cell-targeting signal in *Plasmodium falciparum*-infected erythrocytes. *Proc. Natl Acad. Sci. USA* **100**, 12402–12407 (2003).
14. Marti, M. *et al.* Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. *J. Cell Biol.* **171**, 587–592 (2005).
15. Ansorge, I., Benting, J., Bhakdi, S. & Lingelbach, K. Protein sorting in *Plasmodium falciparum*-infected red blood cells permeabilized with the pore-forming protein streptolysin O. *Biochem. J.* **315**, 307–314 (1996).
16. Mougous, J. D. *et al.* A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
17. Gutensohn, M. *et al.* Toc, Tic, Tet et al.: structure and function of protein transport machineries in chloroplasts. *J. Plant Physiol.* **163**, 333–347 (2006).
18. Sanders, P. *et al.* Identification of protein complexes in detergent-resistant membranes of *Plasmodium falciparum* schizonts. *Mol. Biochem. Parasitol.* **154**, 148–157 (2007).
19. Fischer, K. *et al.* Characterization and cloning of the gene encoding the vacuolar membrane protein EXP-2 from *Plasmodium falciparum. Mol. Biochem. Parasitol.* **92**, 47–57 (1998).
20. Boucher, I. W. *et al.* Structural and biochemical characterization of a mitochondrial peroxiredoxin from *Plasmodium falciparum. Mol. Microbiol.* **61**, 948–959 (2006).
21. Johnson, D. *et al.* Characterization of membrane proteins exported from *Plasmodium falciparum* into the host erythrocyte. *Parasitology* **109**, 1–9 (1994).
22. Freyer, E., Eschenbacher, K. H., Mehlhorn, H. & Rueger, W. Isolation and characterisatin of cDNA clones encoding a 32-kDa dense-granule antigen of *Sarcocystis muris* (Apicomplexa). *Parasit. Res.* **84**, 583–589 (1998).
23. Gilson, P. *et al.* MSP1(19) miniproteins can serve as targets for invasion inhibitory antibodies in *Plasmodium falciparum* provided they contain the correct domains for cell surface trafficking. *Mol. Microbiol.* **68**, 124–138 (2008).
24. Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* **19**, 1794–1802 (2000).
25. Chang, H. H. *et al.* N-terminal processing of proteins exported by malaria parasites. *Mol. Biochem. Parasitol.* **160**, 107–115 (2008).
26. Knuepfer, E. *et al.* Trafficking of the major virulence factor to the surface of transfected *P. falciparum*-infected erythrocytes. *Blood* **105**, 4078–4087 (2005).
27. Boddey, J. A., Moritz, R. L., Simpson, R. J. & Cowman, A. F. Role of the *Plasmodium* Export Element in trafficking parasite proteins to the infected erythrocyte. *Traffic* **10**, 285–299 (2008).
28. Gehde, N. *et al.* Protein unfolding is an essential requirement for transport across the parasitophorous vacuolar membrane of *Plasmodium falciparum. Mol. Microbiol.* **71**, 613–628 (2008).
29. Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243–257 (2001).
30. Eifler, N. *et al.* Cytotoxin ClyA from *Escherichia coli* assembles to a 13-meric pore independent of its redox-state. *EMBO J.* **25**, 2652–2661 (2006).
31. Banumathy, G., Singh, V. & Tatu, U. Host chaperones are recruited in membrane-bound complexes by *Plasmodium falciparum. J. Biol. Chem.* **277**, 3902–3912 (2002).

Author Contributions T.F.d.K.-W. and P.R.G. designed, performed and interpreted much of the experimental work, while B.S.C. designed and interpreted the work and, with T.F.d.K.-W., wrote the manuscript. J.A.B., M.R., P.R.S. and R.J.L. performed experiments and provided intellectual insight in aspects of this study. B.J.S. and A.T.P. contributed the EXP2 molecular modelling and phylogenetic analysis, respectively. A.G.M. and A.F.C. provided novel reagents and mutants, while A.F.C. also provided considerable input into study design and data interpretation. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.S.C. (crabb@burnet.edu.au).

METHODS

Ring-stage proteome. MudPIT (multidimensional protein identification technology) analysis of proteins present in detergent resistant membranes (DRMs) prepared from saponin-lysed ring-stage parasites was performed as described previously[32].

Plasmid construction and transfection. Plasmids pPTEX150-HA/Str3′ and pHSP101-HA/Str3′ were constructed by ligating the last 986 bp of the PTEX150 (PF14_0344) or the last 920 bp of the HSP101 (PF11_0175) coding sequence (excluding the stop codon) into the unique BglII and PstI sites (underlined) of p1.2 (ref. 33). The 3′ targeting sequences were generated by PCR amplification of *P. falciparum* 3D7 genomic DNA using the oligonucleotides PTEX150intF (5′-GAAGATCTGAAGAAGGTATTTTAGATTATGAAG) and PTEX150R-PstI (5′-AACTGCAGCATTGTCGTCCTCTTCTTCGTCC) for PTEX150, and HSP101intF (5′-GCAGATCTCTAAATCCATTATTGGAAATGAAGA) and HSP101intR (5′-AACTGCAGCGGTCTTAGATAAGTTTATAACCAAG) for HSP101. Restriction enzyme sites used for cloning are underlined. This resulted in the fusion of the 3′ end of PTEX150 or HSP101 with the 3×HA/Str epitope tags in p1.2 such they formed a continuous open reading frame. pPTEX150-HA3′ and pHSP101-HA3′ (75 µg) were transfected into *P. falciparum* 3D7 parasites as previously described[34].

Expression of fusion proteins and generation of antibodies. The DNA sequence corresponding to amino acids 181–236 of PTEX150 was amplified from *P. falciparum* 3D7 gDNA using the oligonucleotides PTEX150Ab2F (5′-CGCGGATCCGAAATAAAAATGAAGATGAACTCAC) and PTEX150Ab2R (5′-AACTGCAGATTATCTTTATTCATCATTTTTCC) and ligated into the BamHI and PstI sites (restriction enzyme sites underlined) of a modified version of pMal-c2x (New England Biolabs) that generates a C-terminal 6× His tag[32]. The DNA sequence corresponding to amino acids 68–170 of HSP101 was amplified using the oligonucleotides HSP101e-F (5′-CGCGGATCCCCCGATAATAAGCAAGAGCAAGGA) and HSP101e-R (5′-CCCAAGCTTTTATAATATTGCTTTAATGGCTTCATCGGTT) and ligated into the BamHI and HindIII sites (restriction sites underlined) of pPROEX-HTb (Invitrogen). PTEX150 and HSP101 fusion proteins were expressed in *Escherichia coli* BL21 cells, purified on a Ni-NTA affinity resin (Qiagen) and used to generate monoclonal and polyclonal antibody reagents.

Immunoprecipitations. Mixed ring and trophozoite-stage infected erythrocytes were either treated with 100 U ml^{-1} tetanolysin (Sigma) on ice for 30 min or cross-linked with 2 mM DSP (dithiobis[succinimidyl propionate]) followed by saponin lysis to remove red blood cell material. Pelleted parasites were lysed with 1% (v/v) Triton X-100 in PBS or RIPA buffer (25 mM Tris pH 7.6, 150 mM NaCl, 1% (v/v) Triton X-100, 1% (w/v) Deoxycholate-Na, 0.1% (w/v) SDS) containing complete protease inhibitors (Roche) and lysates then applied to goat anti-HA antibodies coupled to agarose beads (Sapphire BioScience), or mixed with antibodies (10–20 µg) derived against various recombinant malaria proteins for 4 h at 4 °C. In the cases in which uncoupled anti-malaria antibodies were used, Protein G sepharose was then added and the material incubated for a further 4 h at 4 °C. After triplicate washing with 1% (v/v) Triton X-100 in PBS, bound protein was eluted from the beads with 1× SDS–PAGE sample buffer and electrophoresed under reducing conditions. For western blotting, the following dilutions of antibody reagents were used—polyclonal rabbit anti-PTEX150, 1:200; polyclonal rabbit anti-HSP101, 1:1,000; monoclonal mouse anti-HSP101, 1:500; monoclonal mouse anti-EXP2 (mAb 7.7, a gift from J. McBride and described previously[21,35], 1:2,000; monoclonal mouse anti-HA clone 12CA5 (Roche), 1:2,000; polyclonal rabbit anti-MSP(19)[36], 1:1,000; polyclonal rabbit anti-AMA1 (ref. 37), 1:200; monoclonal mouse anti-RESA, 1:2,000[38]; polyclonal rabbit anti-PF08_0137, 1:1,000.

Indirect immunofluorescence analysis (IFA). IFAs were performed on parasites fixed with ice-cold 90% acetone/10% methanol using the following antibody concentrations—polyclonal rabbit anti-PTEX150, 1:100; polyclonal rabbit anti-HSP101, 1:100; monoclonal mouse anti-HSP101, 1:100; monoclonal mouse anti-EXP2 (mAb 7.7),1:20; monoclonal mouse anti-HA, 1:50; polyclonal rabbit anti-MSP(19), 1:1,000. The appropriate Alexa Fluor 488/595 secondary IgG antibodies (Molecular Probes) were used at 1:1,000 dilution. Mounted slides were visualized with a Carl Zeiss Axioskop microscope. Pictures were recorded using a PCO Sensi-Cam and images were produced using ImageJ software (available from the public domain).

Protease sensitivity assays. Transgenic parasite lines (2–5 ml cultures) at ring-trophozoite stage were harvested, washed in PBS and resuspended in either 1HU equinatoxin (5 mg ml^{-1} in PBS) for 6 min at room temperature, 0.09% (w/v) saponin for 10 min on ice or in 1% (v/v) Triton X-100 for 10 min on ice. Parasite samples were then resuspended in either 20 µg ml^{-1} proteinase K (Roche) or in PBS and incubated for 30 min at room temperature. Proteolysis was stopped by the addition of TCA to a final volume of 15% (v/v) and incubation on ice for

15 min. The pellets generated after centrifugation were resuspended in reducing sample buffer for analysis on 4–12% Bis-Tris gels.

Modelling of HSP101. For *P. falciparum* HSP101, homologues of known structure were identified by sequence-to-hidden Markov model comparison in HHpred[39]. Residues 1–26 constitute the ER signal sequence and were omitted. The top two matches, covering *P. falciparum* HSP101 from residue 35 to the C terminus, were to *Thermus thermophilus* ClpB (1QVR) at 38% pairwise sequence identity and to *Escherichia coli* ClpA (1R6B) at 33%. The structure of the protein was modelled in MODELLER[40] with 1QVR, chain B, as a template, except in gapped regions and at the C terminus, where 1R6B was used as a second template. The guiding alignment is shown in Supplementary Fig. 1. The two ATPase rings—one comprising domains N and D1, and the other domain D2—were modelled separately, assembled into hexameric rings using a previously described procedure[41], docked against each other with RosettaDock[42], and connected in Swiss PDB Viewer[43] to obtain the final model. Coordinates for the model can be downloaded via the 'computational models' hyperlink at http://www.eb.tuebingen.mpg.de/departments/1-protein-evolution. Structure figures were rendered in MacPyMOL (DeLano Scientific).

Homology modelling of EXP2. A protein-sequence BLAST search[44] identified 5 sequences homologous to that of EXP2, GI:156099352 from *Plasmodium vivax*, GI:193810225 from *Plasmodium knowlesi*, GI:82596841 from *Plasmodium yoelii*, GI:68069165 from *Plasmodium berghei*, and GI:70943327 from *Plasmodium chabaudi*. These sequences were used to identify potential compatible structures for EXP2 using the FUGUE program[29]. The structure of haemolysin E (PDB code 1QOY) was identified to have the highest level of compatibility with the sequences, albeit marginal (that is, ~90% confidence level). A model of EXP2 using the structure of haemolysin E as a template was constructed using the MODELLER program[40] using the sequence alignment presented in Supplementary Fig. 6A. Twenty-five models of EXP2 were prepared, and the structure with lowest DOPE score[45] was selected for further analysis (Supplementary Fig. 6B). The sequence of EXP2 contains four cysteine residues; cysteine residues 24 and 113 are predicted to be disulphide bonded, whereas cysteine residues 140 and 210 are predicted to lie at the opposite end of the molecule, near the β-sheet structure, separated from one another by ~20 Å. Analysis of the DOPE score for individual residues indicates that the region between residues 220 and 250 is inadequately modelled (the FG-loop and helix-G; Supplementary Fig. 6C), which is the region that is the most different for the sequences in the various species identified in the above BLAST search.

32. Sanders, P. R. *et al.* Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum. J. Biol. Chem.* **280**, 40169–40176 (2005).

33. Baum, J. *et al.* Reticulocyte-binding protein homologue 5 — an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum. Int. J. Parasitol.* **39**, 371–380 (2009).

34. Fidock, D. A. & Wellems, T. E. Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc. Natl Acad. Sci. USA* **94**, 10931–10936 (1997).

35. Hall, R. *et al.* Antigens of the erythrocytes stages of the human malaria parasite *Plasmodium falciparum* detected by monoclonal antibodies. *Mol. Biochem. Parasitol.* **7**, 247–265 (1983).

36. de Koning-Ward, T. F. *et al.* A new rodent model to assess blood stage immunity to the *Plasmodium falciparum* antigen merozoite surface protein 1$_{19}$ reveals a protective role for invasion inhibitory antibodies. *J. Exp. Med.* **198**, 869–875 (2003).

37. Healer, J. *et al.* Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum. Mol. Microbiol.* **52**, 159–168 (2004).

38. Rug, M. *et al.* Correct promoter control is needed for trafficking of the ring-infected erythrocyte surface antigen to the host cytosol in transfected malaria parasites. *Infect. Immun.* **72**, 6095–6105 (2004).

39. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

40. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

41. Diemand, A. V. & Lupas, A. N. Modeling AAA+ ring complexes from monomeric structures. *J. Struct. Biol.* **156**, 230–243 (2006).

42. Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299 (2003).

43. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).

44. Altschul, S. F. *et al.* Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

45. Eramian, D. *et al.* A composite score for predicting errors in protein structure models. *Protein Sci.* **15**, 1653–1666 (2006).

ARTICLES

The *Listeria* transcriptional landscape from saprophytism to virulence

Alejandro Toledo-Arana[1,2,3]†, Olivier Dussurget[1,2,3], Georgios Nikitas[2,4,5], Nina Sesto[1,2,3], Hélène Guet-Revillet[1,2,3], Damien Balestrino[1,2,3], Edmund Loh[6,7], Jonas Gripenland[6,7], Teresa Tiensuu[6,7], Karolis Vaitkevicius[6,7], Mathieu Barthelemy[8], Massimo Vergassola[9,10], Marie-Anne Nahori[1,2,3], Guillaume Soubigou[8], Béatrice Régnault[8], Jean-Yves Coppée[8], Marc Lecuit[2,4,5,11], Jörgen Johansson[6,7] & Pascale Cossart[1,2,3]

The bacterium *Listeria monocytogenes* is ubiquitous in the environment and can lead to severe food-borne infections. It has recently emerged as a multifaceted model in pathogenesis. However, how this bacterium switches from a saprophyte to a pathogen is largely unknown. Here, using tiling arrays and RNAs from wild-type and mutant bacteria grown *in vitro*, *ex vivo* and *in vivo*, we have analysed the transcription of its entire genome. We provide the complete *Listeria* operon map and have uncovered far more diverse types of RNAs than expected: in addition to 50 small RNAs (<500 nucleotides), at least two of which are involved in virulence in mice, we have identified antisense RNAs covering several open-reading frames and long overlapping 5′ and 3′ untranslated regions. We discovered that riboswitches can act as terminators for upstream genes. When *Listeria* reaches the host intestinal lumen, an extensive transcriptional reshaping occurs with a SigB-mediated activation of virulence genes. In contrast, in the blood, PrfA controls transcription of virulence genes. Remarkably, several non-coding RNAs absent in the non-pathogenic species *Listeria innocua* exhibit the same expression patterns as the virulence genes. Together, our data unravel successive and coordinated global transcriptional changes during infection and point to previously unknown regulatory mechanisms in bacteria.

Listeria monocytogenes is responsible for food-borne infections[1]. After ingestion of contaminated food products, bacteria multiply in the intestinal lumen, cross the intestinal barrier and disseminate to deeper organs resulting in brain and materno–fetal infections[2]. Decades of bacterial genetics, cell biology and *in vivo* studies have ranked *Listeria* among the best models to study host–pathogen interactions and intra-cellular parasitism[3]. Determination of the genome sequences of *L. monocytogenes* and of the non-pathogenic *Listeria innocua* paved the way for a full understanding of *L. monocytogenes* pathogenic potential[4,5]. Several genes present in *L. monocytogenes* but absent in *L. innocua* are key virulence factors, such as genes of the 'virulence locus' that encode listeriolysin O, which promotes escape from the vacuole, and ActA, which allows intra- and inter-cellular movements. At least 47 other genes scattered in the genome are associated with virulence.

Coordinated virulence gene expression is critical for infection and orchestrated by quorum sensing or global regulators. PrfA is the main activator of *Listeria* virulence gene expression[6,7]. Its expression is regulated by an RNA thermosensor[8]. Other regulators include SigB, an alternative sigma factor, and VirR, the regulator of a two-component system[9–11]. Small RNAs (sRNAs) have emerged as regulators of virulence in several bacteria. Twenty-one sRNAs have been reported in *L. monocytogenes*[12–15]. None has been described to be involved in virulence[16]. However, a mutant in *hfq*, which encodes an RNA-binding protein, is affected in virulence[17].

To understand how *L. monocytogenes* switches from a saprophyte to a pathogen, we investigated the transcription of its entire genome. Tiling arrays have recently been used for genome-wide expression analysis[18–20]. Two prokaryotic tiling arrays were instrumental in the discovery of transcription start sites and sRNAs in *Escherichia coli* and *Caulobacter crescentus*[21–23] (Supplementary Fig. 1). Here, using a tiling microarray—made of 25-base oligonucleotides with a 9-nucleotide overlap—covering both strands of *L. monocytogenes* chromosome uniformly, we conducted the first unbiased transcriptomic study in a bacterial species and depicted the whole genome transcriptional landscape of *L. monocytogenes* in diverse growth conditions, including *in vivo*.

From the genome to the RNome

We compared the transcripts of EGD-e, the first sequenced strain, grown to exponential phase at 37 °C in rich medium (brain-heart infusion, BHI)—our reference condition—with three physiologically relevant conditions for *L. monocytogenes*: stationary phase, hypoxia and low temperature (30 °C). ΔprfA, ΔsigB and Δhfq mutants grown in BHI to exponential phase were also compared with the reference condition. We then inoculated axenic mice orally with EGD-e, ΔprfA and ΔsigB mutants, and RNAs from bacteria present in the intestinal lumen of mice 24 h post-inoculation were extracted. We also inoculated blood from healthy human donors with EGD-e, ΔprfA or ΔsigB mutants, and bacterial RNAs were extracted after 30 min and 60 min of incubation.

We analysed the annotated open reading frames (ORFs) and found that *Listeria* expresses more than 98% of its ORFs in any condition (Supplementary Fig 2a). Because functionally related genes are generally clustered in operons, identification of operons is critical for gene function elucidation. Here, taking into account all conditions tested, we identified 517 polycistronic operons encoding 1,719 genes which

Figure 1 | *L. monocytogenes* operon map and small RNAs. a, Light and dark blue arrows indicate successive operons revealed by various growth conditions. Purple and orange boxes corresponding to the virulence locus and flagellum biosynthesis operon are detailed at the bottom of the panel. Exp., exponential; Stat., stationary. Colour code bar indicates relative expression. **b,** Genomic map showing the localization of sRNAs of *L. monocytogenes*. Yellow arrows indicate sRNAs absent in *L. innocua*. **c,** sRNA relative expression. Exp., exponential; Stat., stationary. Colour code bar indicates expression fold-change compared to the reference condition. Asterisk, absent in *L. innocua*. WT, wild type. **d,** Examples of transcriptional tiling maps of sRNAs in different conditions. The plots show normalized hybridization intensities (y axis) and genomic coordinates (x axis, in bp). Each dot corresponds to the average of intensity signals from three independent biological repetitions for one probe. Annotated ORFs and sRNAs are shown as blue and orange arrows, respectively. RACE results are shown next to the 1 kb ladder. **e,** Venn diagram for sRNAs. Light blue, *L. monocytogenes* sRNAs; blue, sRNAs absent in *L. innocua*; yellow, sRNAs regulated in intestinal lumen; red, sRNAs regulated in blood; purple, sRNAs regulated in blood and absent in *L. innocua*; green, sRNAs regulated in intestinal lumen and absent in *L. innocua*; orange, sRNAs regulated in intestinal lumen and blood; white, sRNAs regulated in intestinal lumen and blood and absent in *L. innocua*.

represent 60% of the annotated genome (Fig. 1a and Supplementary Table 1). Approximately half of the operons were bicistronic and more than 80% contained less than five cistrons (Supplementary Fig. 2b, c). Thus, our chip design experimentally led, to our knowledge, to the first bacterial operon map.

Listeria small RNAs

Among the 21 sRNAs previously identified (listed 1 to 21 in Fig. 1b, c and Supplementary Table 2), 3 were predicted to act as antisense RNAs and 5 are absent from *L. innocua*[14]. We identified 29 previously unknown sRNAs, the size of which varied from 77 to 534 nucleotides (Supplementary Table 2). Seven are predicted to act as *cis*-antisense. Eight transcripts were mapped using a modified RACE (rapid amplification of cDNA ends), allowing simultaneous determination of 5′- and 3′-ends. In all cases, RACE results correlated with tiling signals (Fig. 1d and Supplementary Fig. 3). Fifteen novel sRNAs (Fig. 1b, c, e) are absent from *L. innocua*. Five (Rli37, Rli40, Rli41, Rli43, Rli44) are predicted to contain small ORFs with appropriated ribosome-binding sites (RBSs; ORF sizes from 28 to 64 amino acids; Supplementary Table 2). Our original annotation of the genome was based on prediction of ORFs larger than 50 amino acids[4]. Thus, only few ORFs were missing in the original annotation.

Several sRNAs were differentially expressed in some conditions (Fig. 1c). Major changes occurred after growth in the intestinal lumen or in blood. PrfA and Hfq did not affect the level of expression of the sRNAs. In contrast, SigB affected expression of several sRNAs.

To predict targets for the previously unknown sRNAs, we used a modified version of the method that we previously described[14]. Instead of scanning regions around the translational start site, we scanned the whole genome of *L. monocytogenes* for RNAs forming duplexes with a given sRNA (Supplementary Table 3). Several sRNAs could interact with other sRNAs, for example, *rli44* could interact with *rli46* and *rli50* with *rli45*, highlighting regulatory networks via interactions between several sRNAs. Most other targets corresponded to genes involved in metabolism. Nevertheless, several sRNAs, *rli28*, *rli32*, *rli34*, *rli38*, *rli49* and *rli50*, paired with messenger RNAs potentially more directly involved in virulence such as internalin family proteins (Supplementary Table 3).

rli38, which is absent in *L. innocua*, showed a ~25-fold increase in blood (Figs 1c and 2a) and high levels when bacteria were grown in the presence of H_2O_2 (data not shown). Moreover, *rli38* could pair to three mRNAs that could have a function in bacterial adaptation to blood, including *fur*, which encodes the global iron uptake regulator (Supplementary Table 3). A *rli38* deletion mutant was generated and, when orally inoculated to mice[24], was attenuated, revealing a role for *rli38* in virulence (Fig. 2a). A deletion was also created in the previously identified *rliB* (ref. 14). In this case also, virulence was affected (Fig. 2b). However, the mutant colonized the liver faster than the wild type, highlighting a role for *rliB* in controlling virulence.

Listeria cis-regulatory RNA elements

Cis-regulatory RNA elements can be long 5′ UTRs (untranslated regions), which through alternative conformations modulate downstream genes expression[25,26], or *cis*-antisense RNAs controlling expression of the mRNA transcribed from the opposite strand.

Conformational changes in 5′ UTRs occur by temperature shift (for example, the *prfA* thermosensor)[8] or after binding a metabolite or a transfer RNA leading to premature transcription termination, arrest in translation initiation or both (for example, riboswitches)[27]. Thus, in certain conditions, the 5′ UTR produces an sRNA. For *L. monocytogenes*, 42 riboswitches are annotated in the Rfam database[28] but none had been experimentally demonstrated besides the PrfA thermosensor[8]. In this study we readily identified 40 riboswitches and distinguished 'on' and 'off' situations in some of the conditions tested (Supplementary Table 2). We focused on the unique lysine riboswitch (LysRS) predicted in *Listeria* (Fig. 2c). Bacteria were grown in minimum medium with or without lysine, and RNAs were

analysed by northern blot (Fig. 2d). In the absence of lysine, in addition to the small transcript corresponding to the riboswitch (LysRS), a transcript comprising the riboswitch and the downstream *lmo0798* gene, which encodes the lysine transporter, was detected. In presence of lysine, the long transcript (LysRS + Lmo0798) was not detected. In contrast, LysRS was even more easily detectable. We could also detect a transcript comprising the *lmo0799* gene and the LysRS, indicating that the riboswitch acts as a gene terminator (Fig. 2c, d). This unexpected observation helped interpretation of why the two cobalamin riboswitches (CobRSs) were not considered as bona fide riboswitches. Indeed, both CobRSs are located in 3′ UTRs and act as transcription terminators (Supplementary Fig. 4b). In one case, there is no proximal downstream gene on the same DNA strand. The predicted riboswitch corresponds to the sRNA named *rli39* and validated by RACE (Supplementary Fig. 3). Four other examples of riboswitches acting as transcription terminators are presented in Supplementary Fig. 4.

Thirteen additional long 5′ UTRs were identified. One corresponds to the 5′ UTR of *mpl*, which encodes the metalloprotease upstream of *actA* and which is required for virulence. With RNA from bacteria grown in BHI or the intestinal lumen, a short transcript (~120 nucleotides) was detected that starts at the *mpl* transcription start site and stops upstream of the *mpl* RBS, highlighting a post-transcriptional mechanism as previously proposed[29]. Because no Rho-independent transcription terminator could be detected, the underlying mechanism is probably not through a riboswitch.

Interestingly, three antisense RNAs cover more than one ORF (Fig. 3a, b). In all cases, northern blots together with the tiling array data identified two different transcripts (RNA1 and RNA2) starting at the same position but with different transcription termination sites (Supplementary Table 4 and Supplementary Figs 5–7). Rho-independent terminators could only be identified for the small transcripts.

Overlapping untranslated regions

We have identified long overlapping 3′ UTRs of convergent genes by detecting nine mRNAs for which transcription ends inside or after the gene located on the opposite strand. The mRNA thus overlaps the mRNA from the convergent gene (from 260 to 1,278 nucleotides of overlap; Supplementary Table 4, Fig. 3c and Supplementary Figs 8–13). How these overlapping 3′ UTRs affect gene expression is unknown.
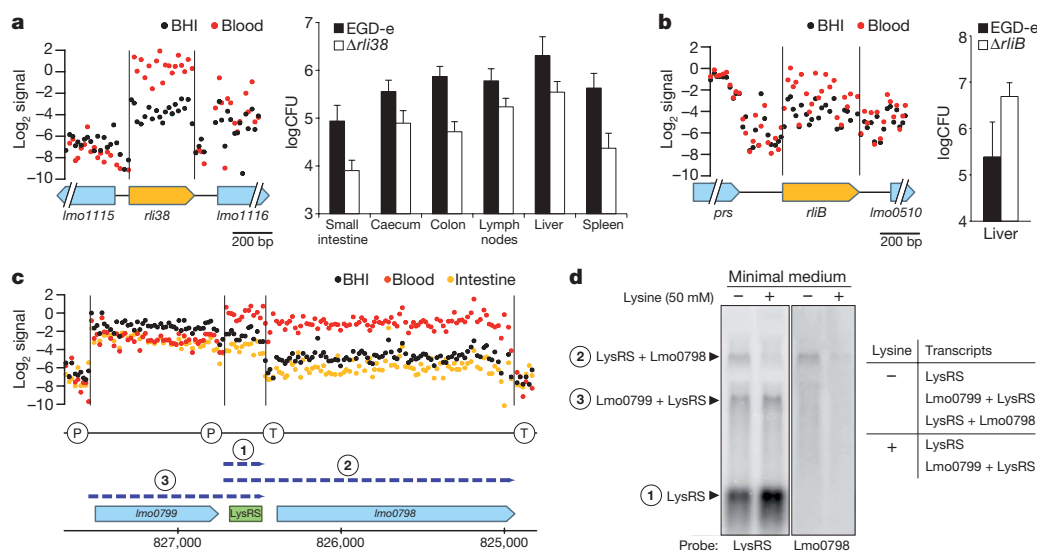
Figure 2 | Functional analysis of Rli38, RliB and LysRS. a, Left: transcriptional tiling map showing expression of *rli38*. Right: bacterial counts of EGD-e ($n = 3$) and Δ*rli38* ($n = 4$) in various organs 72 h after oral infection of humanized E-cadherin knock-in mice[24]. Error bars indicate standard deviations. **b**, Left: transcriptional tiling map showing expression of *rliB*. Right: bacterial counts of EGD-e ($n = 4$) and Δ*rliB* ($n = 4$) in the liver of BALB/c mice 72 h after intravenous inoculation. Error bars indicate standard deviations. **c**, Transcriptional tiling map showing the expression of the LysRS region. P, promoter; T, terminator. **d**, Northern blot after bacterial growth in minimal medium with or without lysine.
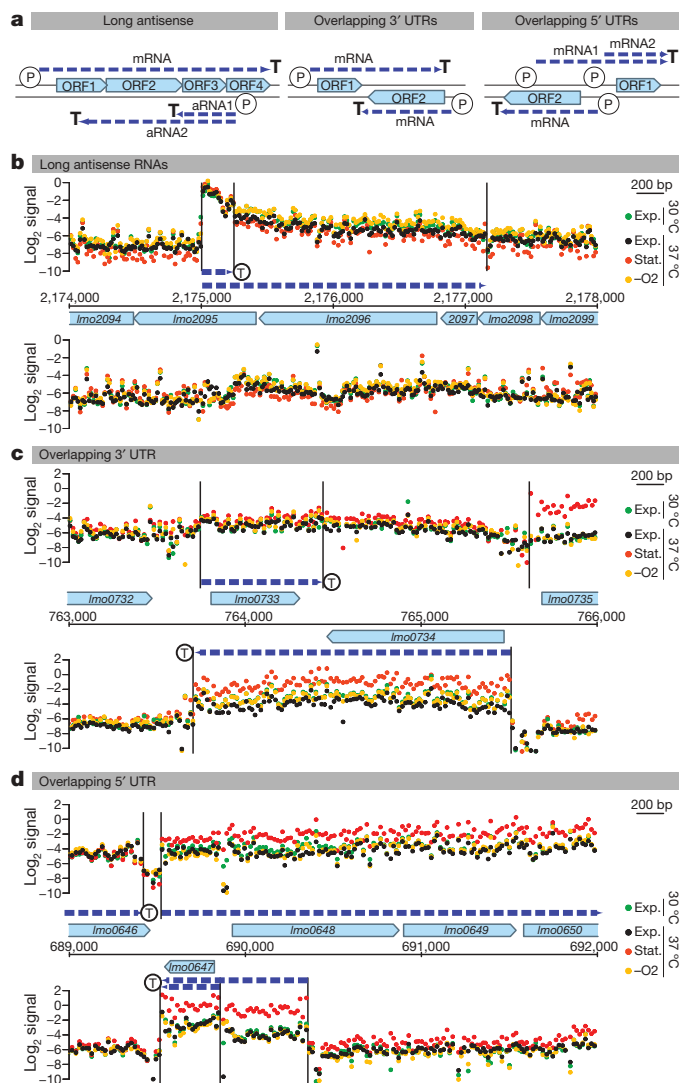
Figure 3 | Cis-regulatory RNA-elements. a, Schematic representation of different types of antisense RNAs found in *L. monocytogenes*. ORFs are shown as blue arrows. Dashed blue lines represent the detected transcripts. P, promoter; T, terminator. Transcriptional tiling maps of a long antisense RNA (**b**), overlapping 3′ UTR of convergent genes (**c**) and overlapping 5′ UTR of divergent genes (**d**) in four growth conditions are shown. Exp., exponential; Stat., stationary. The plots show the normalized hybridization intensities (*y* axis) and genomic coordinates (*x* axis, in bp). Each dot corresponds to the average of intensity signals from three independent biological repetitions for one probe.

We also found four long overlapping 5′ UTRs. In these cases, transcription starts from promoters located inside the ORF of the opposite strand (from 268 to 1,421 nucleotides of overlap; Supplementary Table 4, Fig. 3d and Supplementary Figs 14 and 15). The most unexpected situation concerned the flagellum biosynthesis pathway. In *Listeria*, flagellum synthesis is regulated by temperature, with higher expression at low temperature, and by MogR, a transcriptional repressor of flagellum genes[30]. Here we identified two promoters for MogR (Fig. 4a, b). RACE from circularized RNAs confirmed the presence of two promoters, P1 at 1,697 nucleotides from ATG and P2 at 45 nucleotides from the ATG of MogR, and a common Rho-independent transcription termination (Fig. 4b). A SigB box could be identified upstream of P1. Thus, P1 generates a long 5′ UTR, which overlaps three genes required for the synthesis of the *Listeria* flagellum on the opposite strand, *lmo0675*, *lmo0676* and *lmo0677* (1,421 nucleotides of overlap). The mRNA from P2 is constitutively expressed. Using green fluorescent protein fusions, we showed that P1 is not affected by temperature (Fig. 4c). However,

the P1 transcript is overexpressed in stationary phase and not expressed in a Δ*sigB* mutant as confirmed by northern blot analysis (Fig. 4d). Interestingly, Δ*sigB* mutants in *L. monocytogenes* 10403S and *L. innocua* FSL C2-008 have an increased motility at 30 °C, albeit by an unknown regulatory mechanism[31]. We propose that the increased mobility of the Δ*sigB* mutant is partly due to a lower expression of *mogR*, which in the Δ*sigB* mutant is only expressed from P2. In agreement with this hypothesis, overexpression of the long *mogR* transcript impairs motility (Supplementary Fig. 16a). In addition, expression of the long *mogR* transcript leads to a decrease in the amount of flagellin gene transcripts due to base pairing and processing (data not shown). However, absence of the long *mogR* transcript is not sufficient to explain the phenotype of the Δ*sigB* mutant because point mutations in the SigB box of P1 that totally abrogate transcription from this promoter lead to an increase in
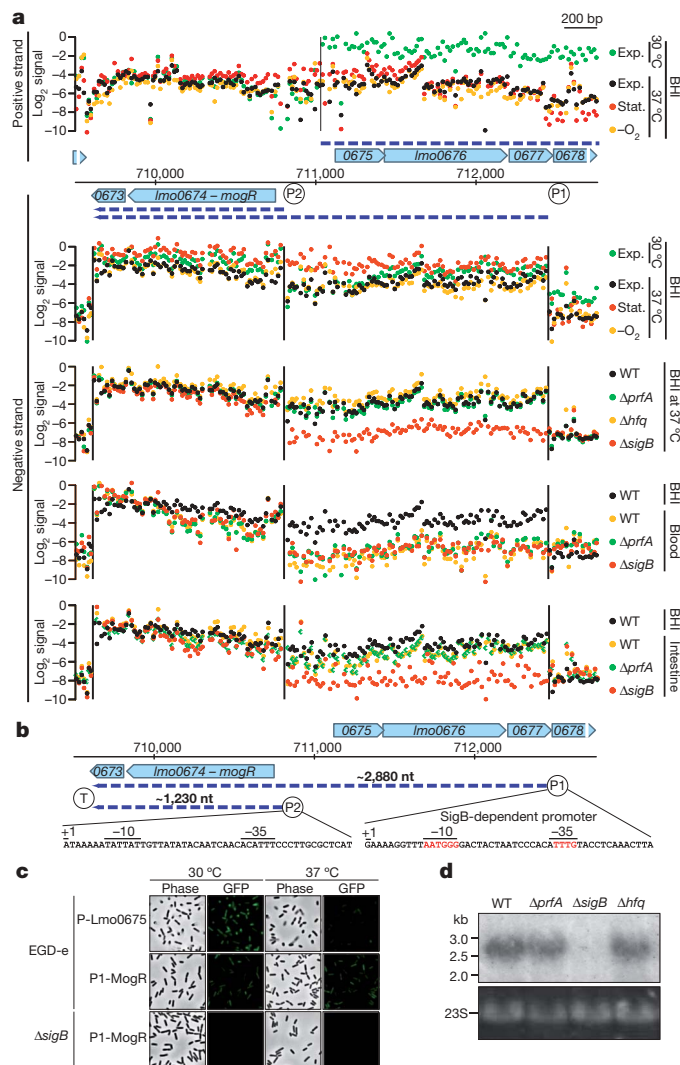
Figure 4 | The MogR locus. a, Transcriptional tiling maps of the *mogR* locus in different conditions and backgrounds. Exp., exponential; Stat., stationary. The plots show the normalized hybridization intensities (*y* axis) and genomic coordinates (*x* axis, in bp). Each dot corresponds to the average of intensity signals from three independent biological repetitions for one probe. Annotated ORFs are shown as blue arrows. Dashed blue lines represent the detected transcripts. **b**, Schematic representation of the two *mogR* promoters validated by RACE, P1 and P2. The consensus SigB binding site is highlighted in red. **c**, Phase contrast and fluorescence microscopy of EGD-e and Δ*sigB* strains expressing GFP under the control of the promoter for *lmo0675* or the P1 promoter of *mogR*, at 30 °C or 37 °C. **d**, Northern blot using an oligonucleotide probe located in the 5′ UTR of the *mogR* transcript.

flagellum gene expression but not to an increase in motility (Supplementary Fig. 16b–d).

Transcriptional switch from saprophytism to virulence

We analysed the *Listeria* protein-coding-gene expression levels using both the tiling arrays and the gene-expression arrays also present on our chip. Because large transcriptional changes were observed, genes were clustered according to similar patterns of expression using the Genesis software (Supplementary Figs 17–48). We then examined whether the sRNAs followed the same pattern of expression.

As shown in Fig. 5, in stationary phase compared to exponential phase, at 37 °C in BHI, 1,016 genes were differentially expressed with a difference higher than twofold (*P* value < 0.05). A total of 490 genes were upregulated (from 2- to 46-fold) including genes described to be involved in virulence and stress responses, for example, *prfA*, *inlA*, *inlB*, *inlH* (*lmo0263*), *ctsR* (*lmo0229*) and *bsh* (*lmo2067*). At 30 °C, 160 genes were upregulated, including 44 genes required for flagellum synthesis (increase of up to 65-fold; Supplementary Fig. 26). Bacteria under hypoxia differentially expressed 161 genes. Approximately 73% of them were downregulated and most were involved in carbohydrate transport and metabolism, whereas upregulated genes included *inlA*, *inlB* (Supplementary Fig. 27), the internalin *lmo0514* (Supplementary Fig. 23) and the peptidoglycan amidase *lmo1216* (Supplementary Fig. 24).

Comparison of gene expression levels in the intestine and in BHI revealed an extensive transcriptional reshaping, with 437 upregulated (from 2- to 29-fold) genes and 769 downregulated (from 2- to 119-fold) genes. Many were related to bacterial stress responses and hypoxic lifestyle. Interestingly, 232 genes were under the control of SigB (Supplementary Fig. 33, clusters 3 and 4, and Supplementary Figs 36 and 37). This number is close to the previously determined SigB regulon[32,33]. Some of these genes encode the invasion proteins, InlA and InlB, and surface proteins, for example, InlH, Lmo0610 and Lmo2085, which could have a role in adhesion to and/or crossing of the intestinal barrier (Fig. 6a, b and Supplementary Fig. 35). Analysis also revealed sigma-B-dependent upregulation of the bile salt hydrolase *bsh* gene required for survival *in vivo*[34] (Supplementary Fig. 37). Together, these results definitively establish the crucial role of SigB for *Listeria* intestinal adaptation as previously proposed[10]. Interestingly, the flagellin gene *flaA*, the cell wall D-alanine-D-alanyl carrier protein ligase gene *dltA* (*lmo0974*), which modifies lipoteichoic acid, and the lipoprotein gene *lpeA* (*lmo1847*) were downregulated in the intestine, indicating that *L. monocytogenes* modifies its surface and appendages in response to intestinal signals (Fig. 6a). *fur* was also downregulated (Fig. 6a). Strikingly, PrfA did not have a significant role in the control of most genes differentially expressed in the intestine (Fig. 6a, b and Supplementary Fig. 33).

In contrast to what was observed in the intestinal lumen, the virulence gene locus was highly overexpressed in blood (Fig. 6a–c), in agreement with the function of these genes, which is to allow escape from the vacuole, intracellular survival and spread inside host cells and tissues. A total of 1,261 genes were differentially expressed in blood: 545 genes showed higher transcript levels whereas 716 showed lower transcript levels (Fig. 5a, b). Transcription of the virulence locus was PrfA-dependent and SigB-independent as reported[35,36] (Fig. 6b and Supplementary Figs 41 and 44). Expression of other virulence gene products such as UhpT, the hexose phosphate transporter[37], internalins InlA and InlB, and the secreted InlC were also induced in blood in a PrfA-dependent but SigB-independent manner (Fig. 6a, b and Supplementary Figs 44 and 45). In both intestine and blood, PrfA- and SigB-independent differential gene expression occurs predicting other regulatory pathways in the coordination of *Listeria* infection (Supplementary Figs 33 and 41).

We examined whether sRNAs would cluster with protein genes. None of the sRNA clustered together with flagellum genes (Supplementary Fig. 17, cluster 9 and Supplementary Fig. 26). Interestingly, *rli33* and *rli38* clustered with genes upregulated in stationary phase and blood
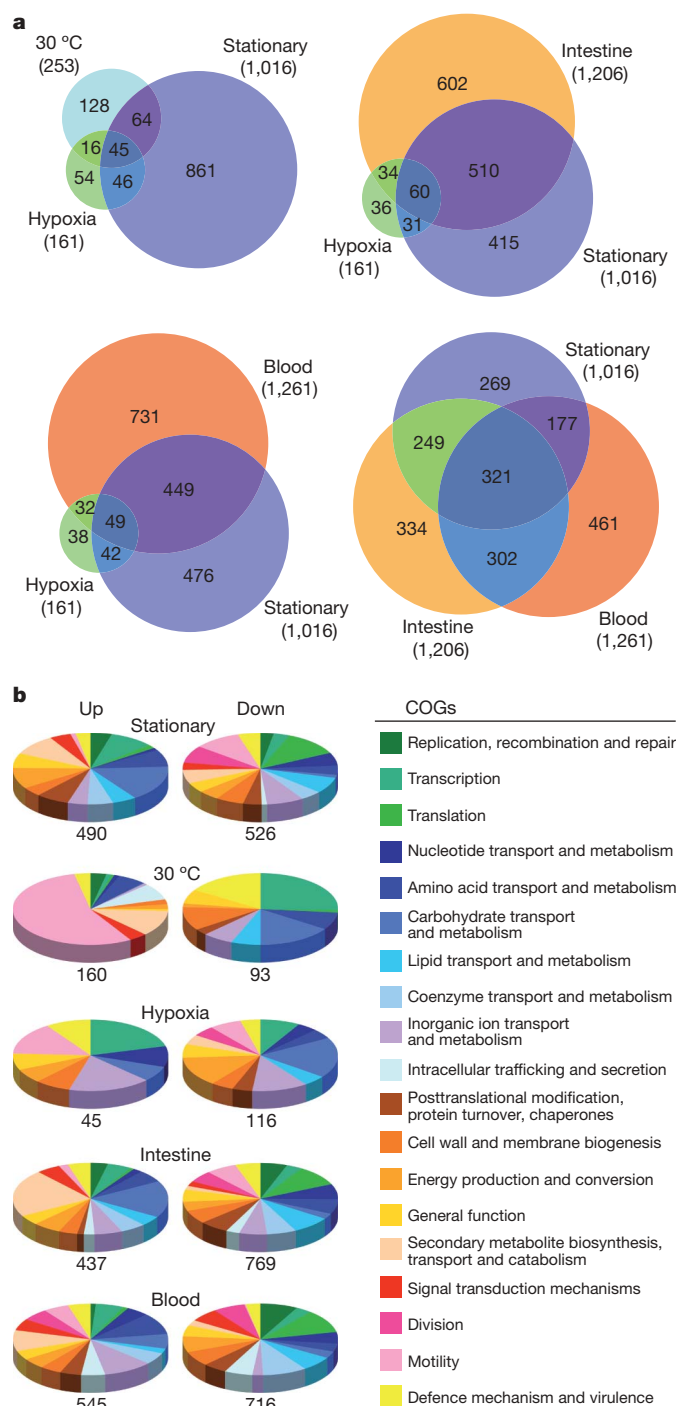
Figure 5 | Gene expression analysis. a, Venn diagrams showing the differentially expressed genes in five conditions. A complete list of the genes can be found in Supplementary Table 6). **b**, Relative abundance of categories of genes differentially expressed by *L. monocytogenes* growing in five growth conditions. Left pie charts, genes upregulated; right pie charts, genes downregulated compared with reference condition. Colours correspond to categories in the COG database.

such as *inlF* (Figs 1c and 6a, d, and Supplementary Fig. 22). *rli28*, *rli46* and *rli47* clustered with genes induced in stationary phase and in intestine (Figs 1c and 6d, and Supplementary Fig. 21). Expression of nine sRNAs, *rli24*, *rli26*, *rli37*, *rli42* and *lhrC1–C5*, was induced in blood and clustered with genes of the virulence gene locus and virulence genes *vip* (*lmo0320*), *uhpT* and *inlC* (Figs 1c and 6d, and Supplementary Fig. 31). These nine sRNAs could have a role during invasive listeriosis after crossing of the intestinal barrier. Finally, several sRNAs, for example, *rli22*, *rli29* and *rli40*, are more expressed in both intestine and blood
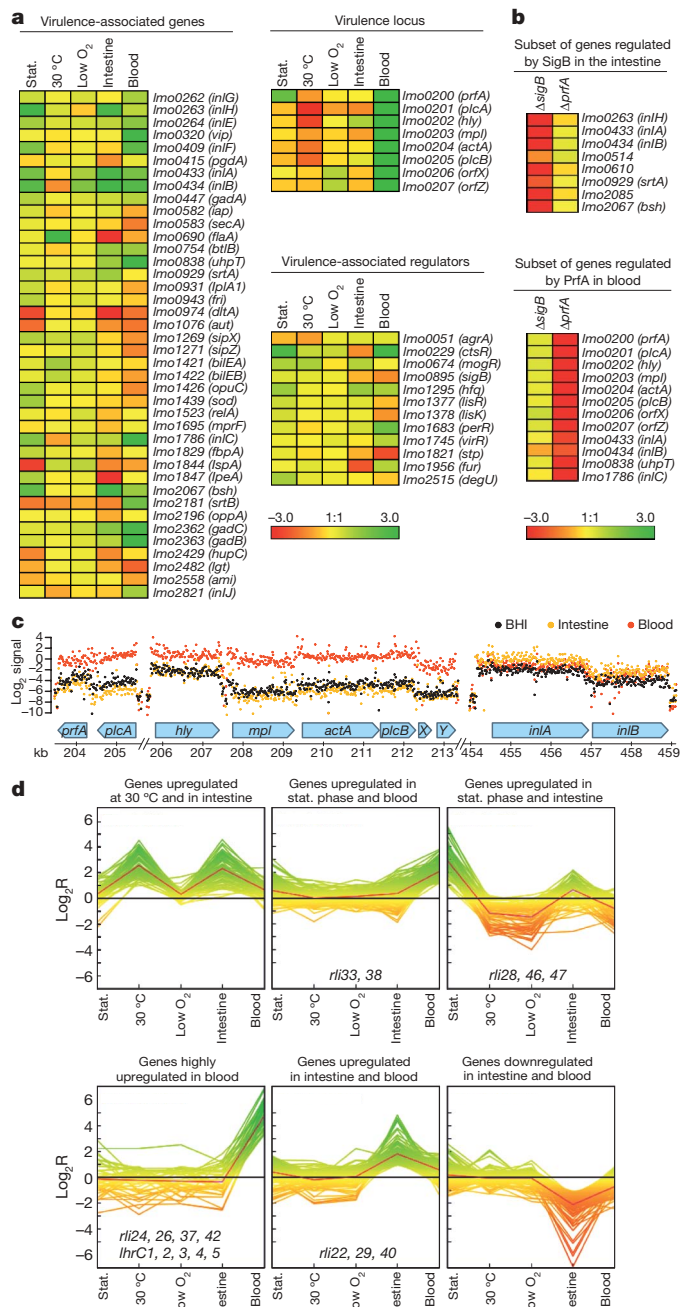
Figure 6 | Virulence gene expression and gene clusters. a, Relative expression of virulence genes in five conditions. Stat., stationary. Colour code bar indicates relative expression compared with reference condition. Usual names given to genes are indicated next to genome-annotated ORFs. **b,** Relative expression of subsets of genes in the ΔsigB and ΔprfA mutants. Colour code bar indicates relative expression compared with EGD-e. The complete list of genes regulated by SigB and PrfA can be found in the Supplementary Information (Supplementary Figs 33–48). **c,** Tiling maps of the virulence locus and the *inlAB* locus. **d,** Clustering of transcripts by their expression scores across five growth conditions compared with the reference condition (exponential growth in BHI at 37 °C). Green to red colours indicate relative expression. The average profile is indicated by a red line. sRNAs with a similar profile are indicated.

(Figs 1c and 6d). They could have a role during switch from saprophytism to intracellular parasitism.

Concluding remarks

We have conducted, to our knowledge, the first comprehensive whole-transcriptome analysis of a bacterial species. We present the first bacterial operon map with identification of 5′- and 3′-end

boundaries in various environments paving the way to a full understanding of the adaptation mechanisms of *Listeria* when the bacterium switches from the environment to the infected host. We highlighted large transcriptional reshaping upon infection and the relevance of SigB in specifically controlling expression of genes important for *L. monocytogenes* adaptation to the intestinal environment, whereas PrfA and the virulence gene cluster are involved in survival and replication in the blood.

Moreover, our study provides evidence that bacterial transcription is more complex than previously anticipated and urges investigations at a single bacterium level. Such studies will no doubt help in unravelling unclear phenotypes encountered when performing deletion analysis or insertion mutagenesis.

METHODS SUMMARY

Bacterial strains and experimental conditions are listed in the Supplementary Methods. RNA was extracted after lysing bacteria in the Fastprep homogenizer (Bio101) using Trizol reagent (Invitrogen). For microarray studies, the *Listeria* tiling array and gene expression array were custom-made by Affymetrix and array hybridization was performed following recommendations of the Affymetrix Genechip expression analysis technical manual. Analysis of tiling-array data was based on the Bioconductor R package. Normalized data were visualized using the Integrated Genome Browser software from Affymetrix. Gene expression analysis was performed using the Genesis software (http://genome.tugraz.at). Mapping of 5′- and 3′-ends of RNAs was performed by RACE using circularized RNAs. The detailed protocol and the oligonucleotides used for RACE are described in the Supplementary Methods and Supplementary Table 5, respectively. Northern blots were carried out using NorthernMax reagents according to the manufacturer's protocol (Ambion). Oligonuceotide probes used for northern blotting are indicated in Supplementary Table 5. For quantitative PCR with reverse transcription, bacterial RNA was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad) or ThermoScript RT–PCR system (Invitrogen) for gene-specific cDNA synthesis. The cDNAs were used as templates for PCR in the presence of SYBR Green supermix (Bio-Rad) in the MyiQ single-colour RT iCycler PCR detection system (Bio-Rad) and data were analysed with the MyiQ optical system software (Bio-Rad). Expression of individual mRNAs was normalized to expression of the *gyrA* gene. Oligonucleotides used for gene amplification are indicated in Supplementary Table 5.

1. Cossart, P. & Toledo-Arana, A. *Listeria monocytogenes*, a unique model in infection biology: an overview. *Microbes Infect.* **10**, 1041–1050 (2008).
2. Lecuit, M. Understanding how *Listeria monocytogenes* targets and crosses host barriers. *Clin. Microbiol. Infect.* **11**, 430–436 (2005).
3. Hamon, M., Bierne, H. & Cossart, P. *Listeria monocytogenes*: a multifaceted model. *Nature Rev. Microbiol.* **4**, 423–434 (2006).
4. Glaser, P. *et al.* Comparative genomics of *Listeria* species. *Science* **294**, 849–852 (2001).
5. Dussurget, O., Pizarro-Cerda, J. & Cossart, P. Molecular determinants of *Listeria monocytogenes* virulence. *Annu. Rev. Microbiol.* **58**, 587–610 (2004).
6. Leimeister-Wachter, M. *et al.* Identification of a gene that positively regulates expression of listeriolysin, the major virulence factor of *Listeria monocytogenes*. *Proc. Natl Acad. Sci. USA* **87**, 8336–8340 (1990).
7. Mengaud, J. *et al.* Pleiotropic control of *Listeria monocytogenes* virulence factors by a gene that is autoregulated. *Mol. Microbiol.* **5**, 2273–2283 (1991).
8. Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110**, 551–561 (2002).
9. Garner, M. R., Njaa, B. L., Wiedmann, M. & Boor, K. J. Sigma B contributes to *Listeria monocytogenes* gastrointestinal infection but not to systemic spread in the guinea pig infection model. *Infect. Immun.* **74**, 876–886 (2006).
10. Gahan, C. G. & Hill, C. Gastrointestinal phase of *Listeria monocytogenes* infection. *J. Appl. Microbiol.* **98**, 1345–1353 (2005).
11. Mandin, P. *et al.* VirR, a response regulator critical for *Listeria monocytogenes* virulence. *Mol. Microbiol.* **57**, 1367–1380 (2005).
12. Barry, T., Kelly, M., Glynn, B. & Peden, J. Molecular cloning and phylogenetic analysis of the small cytoplasmic RNA from *Listeria monocytogenes*. *FEMS Microbiol. Lett.* **173**, 47–53 (1999).
13. Christiansen, J. K. *et al.* Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA* **12**, 1383–1396 (2006).
14. Mandin, P. *et al.* Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.* **35**, 962–974 (2007).
15. Nielsen, J. S. *et al.* Identification of a sigma B-dependent small noncoding RNA in *Listeria monocytogenes*. *J. Bacteriol.* **190**, 6264–6270 (2008).
16. Toledo-Arana, A., Repoila, F. & Cossart, P. Small noncoding RNAs controlling pathogenesis. *Curr. Opin. Microbiol.* **10**, 182–188 (2007).

17. Christiansen, J. K. *et al.* The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J. Bacteriol.* **186**, 3355–3362 (2004).
18. Yazaki, J., Gregory, B. D. & Ecker, J. R. Mapping the genome landscape using tiling array technology. *Curr. Opin. Plant Biol.* **10**, 534–542 (2007).
19. Gregory, B. D., Yazaki, J. & Ecker, J. R. Utilizing tiling microarrays for whole-genome analysis in plants. *Plant J.* **53**, 636–644 (2008).
20. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103**, 5320–5325 (2006).
21. Selinger, D. W. *et al.* RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nature Biotechnol.* **18**, 1262–1268 (2000).
22. McGrath, P. T. *et al.* High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nature Biotechnol.* **25**, 584–592 (2007).
23. Landt, S. G. *et al.* Small non-coding RNAs in *Caulobacter crescentus. Mol. Microbiol.* **68**, 600–614 (2008).
24. Disson, O. *et al.* Conjugated action of two species-specific invasion proteins for fetoplacental listeriosis. *Nature* **455**, 1114–1118 (2008).
25. Winkler, W. C. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* **9**, 594–602 (2005).
26. Brantl, S. Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr. Opin. Microbiol.* **10**, 102–109 (2007).
27. Coppins, R. L., Hall, K. B. & Groisman, E. A. The intricate world of riboswitches. *Curr. Opin. Microbiol.* **10**, 176–181 (2007).
28. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33** (Database issue), D121–D124 (2005).
29. Loh, E., Gripenland, J. & Johansson, J. Control of *Listeria monocytogenes* virulence by 5′-untranslated RNA. *Trends Microbiol.* **14**, 294–298 (2006).
30. Grundling, A., Burrack, L. S., Bouwer, H. G. & Higgins, D. E. *Listeria monocytogenes* regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence. *Proc. Natl Acad. Sci. USA* **101**, 12318–12323 (2004).
31. Raengpradub, S., Wiedmann, M. & Boor, K. J. Comparative analysis of the sigma B-dependent stress responses in *Listeria monocytogenes* and *Listeria innocua* strains exposed to selected stress conditions. *Appl. Environ. Microbiol.* **74**, 158–171 (2008).
32. Kazmierczak, M. J., Mithoe, S. C., Boor, K. J. & Wiedmann, M. *Listeria monocytogenes* sigma B regulates stress response and virulence functions. *J. Bacteriol.* **185**, 5722–5734 (2003).
33. Hain, T. *et al.* Temporal transcriptomic analysis of the *Listeria monocytogenes* EGD-e sigma B regulon. *BMC Microbiol.* **8**, 20 (2008).
34. Dussurget, O. *et al. Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol. Microbiol.* **45**, 1095–1106 (2002).
35. Chakraborty, T., Hain, T. & Domann, E. Genome organization and the evolution of the virulence gene locus in *Listeria* species. *Int. J. Med. Microbiol.* **290**, 167–174 (2000).
36. Scortti, M. *et al.* The PrfA virulence regulon. *Microbes Infect.* **9**, 1196–1207 (2007).
37. Chico-Calero, I. *et al.* Hpt, a bacterial homolog of the microsomal glucose-6-phosphate translocase, mediates rapid intracellular proliferation in *Listeria. Proc. Natl Acad. Sci. USA* **99**, 431–436 (2002).

LETTERS

Strong tidal dissipation in Io and Jupiter from astrometric observations

Valéry Lainey[1], Jean-Eudes Arlot[1], Özgür Karatekin[2] & Tim Van Hoolst[2]

Io is the volcanically most active body in the Solar System and has a large surface heat flux[1-3]. The geological activity is thought to be the result of tides raised by Jupiter[4], but it is not known whether the current tidal heat production is sufficiently high to generate the observed surface heat flow[5,6]. Io's tidal heat comes from the orbital energy of the Io–Jupiter system (resulting in orbital acceleration), whereas dissipation of energy in Jupiter causes Io's orbital motion to decelerate. Here we report a determination of the tidal dissipation in Io and Jupiter through its effect on the orbital motions of the Galilean moons. Our results show that the rate of internal energy dissipation in Io ($k_2/Q = 0.015 \pm 0.003$, where k_2 is the Love number and Q is the quality factor) is in good agreement with the observed surface heat flow[5,6], and suggest that Io is close to thermal equilibrium. Dissipation in Jupiter ($k_2/Q = (1.102 \pm 0.203) \times 10^{-5}$) is close to the upper bound of its average value expected from the long-term evolution of the system[7], and dissipation in extrasolar planets may be higher than presently assumed[8]. The measured secular accelerations indicate that Io is evolving inwards, towards Jupiter, and that the three innermost Galilean moons (Io, Europa and Ganymede) are evolving out of the exact Laplace resonance.

The orbital evolution of the Galilean system due to tidal dissipation can be determined from astrometrically observed positions of the Galilean satellites over an extended period of time[9,10] by using an accurate model of the orbital motion. Most orbital models are based on elaborate analytical methods and include the complex dynamics induced by the interactions of the Galilean moons. In particular, because the three innermost Galilean moons are in the Laplace resonance characterized by the relation

$$L_1 - 3L_2 + 2L_3 \approx \pi$$

where L_1, L_2 and L_3 denote the mean longitudes of Io, Europa and Ganymede, respectively, changes in orbital energy and angular momentum are distributed between the three moons. Unfortunately, some long-period terms are lacking in such models, which explains why previous studies show large and widely different, even contradictory, results (see Table 1 and Supplementary Information for details). Here we use a method that numerically integrates the full equations of motion (Supplementary Information) for the satellite centres of mass instead of using an approximate analytical solution. Moreover, the tidal effects are directly included in the orbital model through the appearance of the Love number, k_2, and the quality factor, Q, in the combination k_2/Q for Io and Jupiter. The orbital effects due to the dissipation in the Galilean satellites other than Io are neglected, but we take into account the tidal bulges raised by each moon on Jupiter using a constant Jupiter quality factor (Supplementary Information).

An extensive set of astrometric observations starting in 1891 and ending in 2007 has been considered in the fitting process. A long,

detailed set of observations such as this is necessary to reveal the long-term effects of dissipation on the orbits. We use a weighted least-squares inversion procedure and minimize the differences between the observed and computed positions of the satellites to determine the parameters of the model, in particular the respective dissipation ratios, k_2/Q, of Io and Jupiter. Our solution for the tidal dissipation yields $k_2/Q = 0.015 \pm 0.003$ (formal error bar, 1σ) for Io and $k_2/Q = (1.102 \pm 0.203) \times 10^{-5}$ for Jupiter. The Io–Jupiter interaction dominates the orbital evolution and is responsible for a large correlation coefficient of 0.983 between the dissipation ratios of Io and Jupiter. The almost 2% difference with respect to unity is due to the evolution of the Laplace resonance and is sufficient to separate the dissipation in Io from that in Jupiter (see Supplementary Information for a complete analysis of this correlation).

The dissipation values correspond to orbital acceleration values, $\dot{n}/n$ (a dot denoting differentiation with respect to time), of $(0.14 \pm 0.01) \times 10^{-10}\,\mathrm{yr}^{-1}$, $-(0.43 \pm 0.10) \times 10^{-10}\,\mathrm{yr}^{-1}$ and $-(1.57 \pm 0.27) \times 10^{-10}\,\mathrm{yr}^{-1}$ (formal error bars, 1σ) for Io, Europa and Ganymede, respectively. These accelerations represent a shift in the satellite orbital positions of respectively 55 km, -125 km and -365 km over the 116 years considered. Surprisingly, the most external moon Ganymede shows the largest drift, as a consequence of the Laplace resonance. The 1σ post-fit astrometric residuals range essentially between 0.02 and 0.15 arcsec (Fig. 1, Supplementary Table 2 and Supplementary Table 3), which corresponds to 60 to 450 km at the distance of Jupiter. Owing to the long time span considered, such accuracy is enough to allow the tidal accelerations to be discerned from the observations.

Table 1 | A selection of secular mean-motion accelerations published for the three inner Galilean moons

Ref.	Secular mean-motion acceleration ($\dot{n}/n$) ($10^{-10}\,\mathrm{yr}^{-1}$)		
	Io	Europa	Ganymede
9	+3.3 ± 0.5	+2.7 ± 0.7	+1.5 ± 0.6
10	−0.074 ± 0.087	−0.082 ± 0.097	−0.098 ± 0.153
24	+4.54 ± 0.95	+5.6 ± 5.7	+2.8 ± 2.0
25	+2.27 ± 0.70	−0.67 ± 0.80	+1.06 ± 1.00
26	+3.6 ± 1.0	—	—
This paper	+0.14 ± 0.01	−0.43 ± 0.10	−1.57 ± 0.27

Refs 9, 24 used a simple orbital model, whereas refs 10, 25, 26 used the much more accurate Sampson–Lieske theory. Nevertheless, this orbital model has internal errors on the order of a few hundred kilometres (Supplementary Information), explaining the lack of agreement between all acceleration values. In particular, most of the tidal acceleration values found (see also Supplementary Table 1) are quite large except for those of ref. 10, in which data from over a very long time span (old eclipses) were used, partly averaging the missing long-period terms. Because our dynamical model fits the k_2/Q ratios, our solution uncertainties in $\dot{n}$ have been derived for each satellite from the comparison of the two simulations that produced the minimum and maximum values of k_2/Q, respectively. As the k_2/Q ratios of Io and Jupiter are highly correlated, we could assume the minimum bound for Io's ratio when introducing the minimum bound of Jupiter's ratio, and vice versa.

[1]IMCCE-Observatoire de Paris, UMR 8028 du CNRS, 77 Avenue Denfert-Rochereau, 75014 Paris, France. [2]Royal Observatory of Belgium, Avenue Circulaire 3, Uccle, 1180 Bruxelles, Belgium.
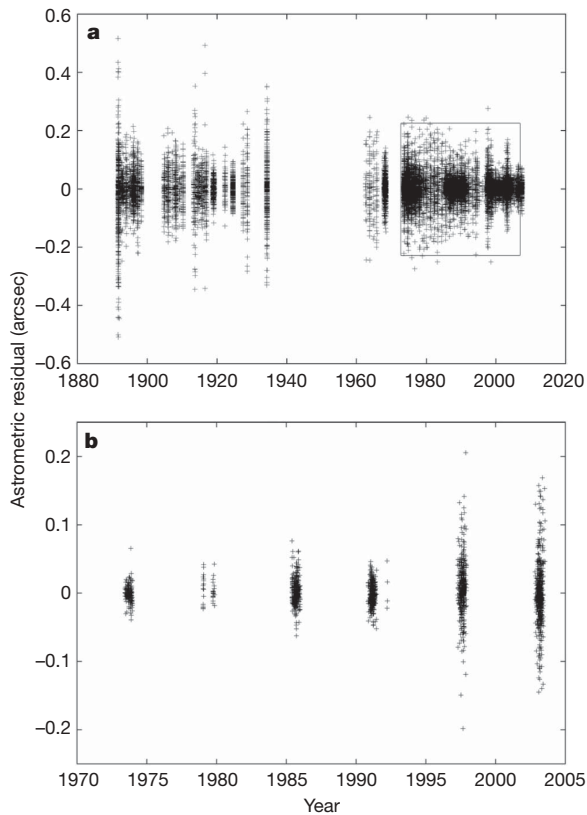
Figure 1 | Astrometric residuals. Residuals between the astrometric observations and our numerical model, after fitting the initial state vectors of each moon and the k_2/Q ratios of Jupiter and Io. We used an extensive set of astrometric observations that started in 1891, with heliometer measurements and the first photographic plates, and continued until 2007, with the most recent observations from the FASTT survey[27]. **b**, Observations of the mutual events (occultation or eclipse of one satellite by another occurring every six years) from 1973 to 2003 (Supplementary Information; boxed region in **a**). The global 1σ accuracy is better than 0.1 arcsec (Supplementary Tables 2 and 3) at the Jovian distance (1 arcsec corresponds to about 3,000 km). The observations of mutual events, known to be among the most accurate observations, have a 1σ accuracy of about 0.025 arcsec and provide the best constraint of the satellite orbits for the past decades. Moreover, instead of considering the position of each satellite as given on the celestial sphere (that is, in arcseconds), we have considered only the relative positions of the observed satellites in the fitting process. For example, for the declination coordinate δ_i we did not fit $\delta_i^{\text{observed}} - \delta_i^{\text{computed}}$ but instead the quantity $[(\delta_i/\delta')^{\text{observed}} - (\delta_i/\delta')^{\text{computed}}]\delta'^{\text{computed}}$, where $\delta' = \sqrt{\sum_{j=1}^{N} \delta_j^2}$ and N is the number of Galilean moons present in an observation. This allowed us to remove systematic errors in the scale factor introduced by the observers during the astrometric calibration of their observations.

The global energy dissipation, $\dot{E}$, in Io can be determined from $k_2/Q = 0.015 \pm 0.003$ using[11]

$$\dot{E} = -\frac{21}{2}\frac{k_2}{Q}\frac{n^5 R^5}{G} e^2$$

where R is the radius of Io, G the gravitational constant and e the orbital eccentricity. We obtain $\dot{E} = (9.33 \pm 1.87) \times 10^{13}$ W. If energy were transported out of Io at the same rate, the associated surface heat flux would be 2.24 ± 0.45 W m^{-2}, which is similar to the observed surface heat flux (Fig. 2). This suggests that Io's interior is close to thermal equilibrium and that Io's internal heat is not the remnant of a past highly dissipative orbital configuration[12–15].

Theoretical studies have not been able to reproduce in a consistent equilibrium model both the tidal energy dissipation and the transport of this internally generated energy to the surface by mantle convection at the observed high surface heat flux[15,16]. It has been argued that mantle viscosities cannot be chosen to satisfy both
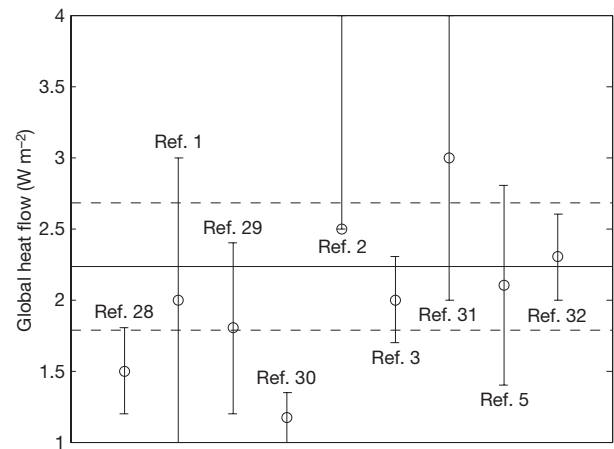
Figure 2 | Comparison of Io's thermal emission with the global dissipation determined in the present study. Io's intensive volcanic activity is associated with a large surface heat flow. The value determined by the present study (2.24 ± 0.45 W m^{-2}, shown by the horizontal lines) is in good agreement with the results of the remote observations[1–3,5,28–32] of Io's thermal emission (the stated error bars mostly indicate ranges for the solutions rather than the standard deviations), suggesting that Io is close to thermal equilibrium. The average surface heat flow of Io is much larger than that of the Earth[33] (0.09 W m^{-2}) and could be comparable to the high surface heat flow of the early Earth.

constraints: viscosities needed to generate the observed tidal dissipation (on the order of 10^{13}–10^{16} Pa s; ref. 11) are too high to transport the produced heat to the surface by convection[16]. If Io is in thermal equilibrium as suggested here, a more efficient heat transport mechanism with a different viscosity dependence on temperature is required. The magma migration in Io's partially molten interior indicated by high eruption temperatures (larger than 1,300 K)[17] is a possible mechanism[18].

Plausible ranges for the quality factor of Io can be studied from the measured ratios k_2/Q by using limit values of the Love number. An upper limit of $Q = 82$ is found by considering Io as a body without strength, that is, by replacing the tidal Love number (k_2) by the fluid Love number, $k_2^{\text{f}} = 1.23$ (ref. 19). A much smaller quality factor is obtained for more realistic models in which Io is composed of a metallic core, a viscoelastic silicate mantle and an elastic lithosphere, depending on the viscosity and shear modulus of the mantle[11]. A typical model with a core radius of 700 km, a core density of 6,944 kg m^{-3}, a mantle density of 3,375 kg m^{-3} and a 40-km-thick lithosphere with a density of 2,600 kg m^{-3}, in agreement with the observations made by NASA's Galileo spacecraft[19], yields $k_2 \approx 0.04$ and $Q \approx 3$ for a mantle rigidity of 50 GPa and viscosity of 4.1×10^{15} Pa s. However, if a low-viscosity asthenosphere exists in Io, the tidal Love number would increase to 0.7–0.8, resulting in a larger quality factor, of ~50.

The dissipation in Jupiter (at the induced frequency of Io's tidal excitation) is determined to be $Q = (3.56 \pm 0.66) \times 10^4$ for the conventional value $k_2 = 0.379$ (ref. 20). This dissipation value is close to the lower limit over the age of the Solar System, $Q \gtrsim 6 \times 10^4$, determined from the expansion of the orbits[7], implying important dissipation in Jupiter. Estimates from numerical dissipation models of Jupiter suggest that Q undergoes large fluctuations as a function of tidal frequency[21,22] and that the observed dissipation may be different from its long-term average (Supplementary Information).

Our results show that the mean motion of Io increases whereas those of Europa and Ganymede decrease. Therefore, Io moves inwards, towards Jupiter, and loses more orbital energy by dissipation of solid-body tides raised by Jupiter and by the Laplace resonance interaction than it gains from the exchange of angular momentum with Jupiter's rotational energy through tidal dissipation in Jupiter. The evolution of the Laplace resonance can be expressed in terms of the rate of change of the satellite mean motions, $\dot{n}_1$, $\dot{n}_2$ and

$\dot{n}_3$, where the subscripts correspond to Io, Europa and Ganymede, respectively[7,13,23]. We have

$$\dot{v} = \dot{n}_1 - 2\dot{n}_2 = \dot{n}_2 - 2\dot{n}_3$$

The system is in stable equilibrium for $\dot{v} = 0$, which requires a balance between the dissipation rates in Jupiter and Io. With our rates of change of the mean motions, we have $\dot{v} = (0.74 \pm 0.24) \times 10^{-7}$ rad yr^{-2}, suggesting that the satellites are evolving away from exact resonance.

1. Matson, D. L., Ransford, G. A. & Johnson, T. V. Heat flow from Io. *J. Geophys. Res.* **86**, 1664–1672 (1981).
2. Veeder, G., Matson, D., Johnson, T., Blaney, D. & Goguen, J. Io's heat flow from infrared radiometry: 1983–1993. *J. Geophys. Res.* **99**, 17095–17162 (1994).
3. Spencer, J. R. *et al.* Io's thermal emission from the Galileo photopolarimeter-radiometer. *Science* **288**, 1198–1201 (2000).
4. Peale, S. J., Cassen, P. & Reynolds, R. T. Melting of Io by tidal dissipation. *Science* **203**, 892–894 (1979).
5. McEwen, A. S., Keszthelyi, P., Lopes, R., Schenk, P. M. & Spencer, J. R. in *Jupiter: The Planet, Satellites and Magnetosphere* (eds Bagenal, F., Dowling, T. E. & McKinnon, W. B.) 307–328 (Cambridge Univ. Press, 2004).
6. Moore, B., Schubert, G., Anderson, J. D. & Spencer, J. R. in *Io After Galileo: A New View of Jupiter's Volcanic Moon* (eds Lopes, R. M. C. & Spencer, J. R.) 89–105 (Springer, 2007).
7. Yoder, C. F. & Peale, S. The tides of Io. *Icarus* **47**, 1–35 (1981).
8. Levrard, B. *et al.* Tidal dissipation within hot Jupiters: a new appraisal. *Astron. Astrophys.* **462**, L5–L8 (2007).
9. De Sitter, W. Orbital elements determining the longitudes of Jupiter's satellites, derived from observations. *Leiden Ann.* **16**(2), 1–92 (1928).
10. Lieske, J. H. Galilean satellite evolution - observational evidence for secular changes in mean motions. *Astron. Astrophys.* **176**, 146–158 (1987).
11. Segatz, M., Spohn, T., Ross, M. N. & Schubert, G. Tidal dissipation, surface heat flow, and figure of viscoelastic models of Io. *Icarus* **75**, 187–206 (1988).
12. Ojakangas, G. & Stevenson, D. Episodic volcanism of tidally heated satellites with application to Io. *Icarus* **66**, 341–358 (1986).
13. Greenberg, R. Galilean satellites - evolutionary paths in deep resonance. *Icarus* **70**, 334–347 (1987).
14. Fischer, H. J. & Spohn, T. Thermal-orbital histories of viscoelastic models of Io (J1). *Icarus* **83**, 39–65 (1990).
15. Hussmann, H. & Spohn, T. Thermal-orbital evolution of Io and Europa. *Icarus* **171**, 391–410 (2004).
16. Moore, W. Tidal heating and convection in Io. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001943 (2003).
17. Keszthelyi, L. *et al.* New estimates for Io eruption temperatures: implications for the interior. *Icarus* **192**, 491–502 (2007).
18. Moore, W. The thermal state of Io. *Icarus* **154**, 548–550 (2001).
19. Anderson, J. D., Jacobson, R. A. & Lau, E. L. Io's gravity field and interior structure. *J. Geophys. Res.* **106**, 32963–32969 (2001).
20. Gavrilov, S. V. & Zharkov, V. N. Love numbers of the giant planets. *Icarus* **32**, 443–449 (1977).
21. Ogilvie, G. I. & Lin, D. N. C. Tidal dissipation in rotating giant planets. *Astrophys. J.* **610**, 477–509 (2004).
22. Wu, Y. Origin of tidal dissipation in Jupiter. II. The value of Q. *Astrophys. J.* **635**, 688–710 (2005).
23. Yoder, C. F. How tidal heating in Io drives the Galilean orbital resonance locks. *Nature* **279**, 747–770 (1979).
24. Goldstein, S. J Jr & Jacobs, K. C. A recalculation of the secular acceleration of Io. *Astron. J.* **110**, 3054–3057 (1995).
25. Vasundhara, R., Arlot, J.-E. & Descamps, P. in *Dynamics, Ephemerides and Astrometry of the Solar System* (eds Ferraz-Mello, S., Morando, B. & Arlot, J.-E.) 145–149 (Proc. 172nd Symp. Int. Astron. Union, IAU, 1996).
26. Aksnes, K. & Franklin, F. A. Secular acceleration of Io derived from mutual satellite events. *Astron. J.* **122**, 2734–2739 (2001).
27. Stone, R. C. Positions for the outer planets and many of their satellites. V. FASTT observations taken in 2000–2001. *Astron. J.* **122**, 2723–2733 (2001).
28. Morrison, D. & Telesco, C. M. Io - observational constraints on internal energy and thermophysics of the surface. *Icarus* **44**, 226–233 (1980).
29. Sinton, W. M. The thermal emission spectrum of Io and a determination of the heat flux from its hot spots. *J. Geophys. Res.* **86**, 3122–3128 (1981).
30. Johnson, T. V. *et al.* Volcanic hotspots on Io - stability and longitudinal distribution. *Science* **226**, 134–137 (1984).
31. Veeder, G. J., Matson, D. L., Johnson, T. V., Davies, A. G. & Blaney, D. L. The polar contribution to the heat flow of Io. *Icarus* **169**, 264–270 (2004).
32. Rathbun, J. A. *et al.* Mapping of Io's thermal radiation by the Galileo photopolarimeter-radiometer (PPR) instrument. *Icarus* **169**, 127–139 (2004).
33. Lay, T., Hernlund, J. & Bussett, B. A. Core–mantle boundary heat flow. *Nature Geosci.* **1**, 25–32 (2008).

LETTERS

Nanomechanical measurements of a superconducting qubit

M. D. LaHaye[1], J. Suh[1], P. M. Echternach[3], K. C. Schwab[2] & M. L. Roukes[1]

The observation of the quantum states of motion of a macroscopic mechanical structure remains an open challenge in quantum-state preparation and measurement. One approach that has received extensive theoretical attention[1–13] is the integration of superconducting qubits as control and detection elements in nanoelectromechanical systems (NEMS). Here we report measurements of a NEMS resonator coupled to a superconducting qubit, a Cooper-pair box. We demonstrate that the coupling results in a dispersive shift of the nanomechanical frequency that is the mechanical analogue of the 'single-atom index effect'[14] experienced by electromagnetic resonators in cavity quantum electrodynamics. The large magnitude of the dispersive interaction allows us to perform NEMS-based spectroscopy of the superconducting qubit, and enables observation of Landau–Zener interference effects—a demonstration of nanomechanical read-out of quantum interference.

Dispersive frequency shifts resulting from the non-resonant interaction of a single atom and a macroscopic photon cavity were first demonstrated over 20 years ago[15], and ultimately have enabled beautiful demonstrations of the quantum nature of light and investigations of quantum decoherence[14]. Some of the most impressive of such measurements include the non-destructive observation of individual microwave photons[16] and the preparation of 'Schrödinger-cat' states of a single cavity mode[17]. Similar effects in superconducting qubits have also been used to detect the Fock states of a coplanar waveguide resonator[18] and the dressed-states of a microwave-driven Cooper-pair box (CPB) qubit[19].

It has been appreciated for some time that a nanomechanical resonator coupled to a superconducting qubit should be formally identical to cavity quantum electrodynamics (CQED) systems, such as a simple harmonic oscillator coupled to a two-level quantum system[1–13]. Furthermore, because of the large frequency difference between typical superconducting qubits and NEMS, a coupling regime that is analogous to the dispersive limit of CQED should exist naturally and, in a similar manner, enable the preparation and measurement of highly non-classical nanomechanical entangled states[6,11–13] and Fock states[2,7–9,11]. In this work, as a first step in implementing these more advanced proposals, we realize dispersive coupling of a CPB qubit and a nanomechanical resonator, and demonstrate, through measurements of the nanoresonator's CPB-state-dependent frequency shift, that the interaction is consistent with the simple picture of a harmonic oscillator coupled to a two-level quantum system.

Our nanomechanical resonator is the fundamental in-plane flexural mode of a suspended silicon nitride nanostructure (Fig. 1a). Its fundamental-mode response can be well described as a damped simple harmonic oscillator with characteristic resonant frequency $\omega_{NR}/2\pi = 58$ MHz (Fig. 1c), effective mass $M \approx 4 \times 10^{-16}$ kg, spring constant $K = M\omega_{NR}^2 \approx 60$ N m^{-1} and damping rate $\kappa = \omega_{NR}/Q$, where Q ranges between ∼30,000 and ∼60,000 (Fig. 1c), depending on the temperature and the resonator's coupling to the measurement circuit and the CPB. Similar to the case for an electromagnetic oscillator, a Hamiltonian operator for the nanoresonator can be written in terms of creation, $\hat{a}^\dagger$, and annihilation, $\hat{a}$, operators, yielding $\hat{H}_{NR} = \hbar\omega_{NR}(\hat{a}^\dagger\hat{a} + 1/2)$, where $\hbar = h/2\pi$ is the reduced Planck constant and the quanta in the mode, of which there are $N = \langle\hat{a}^\dagger\hat{a}\rangle$, are now mechanical quanta.

A split-junction CPB qubit[20], formed from two Josephson tunnel junctions and a superconducting aluminium loop, is coupled to the nanoresonator through capacitance, C_{NR} (Fig. 1a). The CPB is well described by a simple spin-1/2 Hamiltonian[21], $\hat{H}_{CPB} = (E_{el}\hat{\sigma}_z - E_J\hat{\sigma}_x)/2$, where $\hat{\sigma}_z$ and $\hat{\sigma}_x$ are Pauli matrices in the CPB's charge basis. The first term in $\hat{H}_{CPB}$ is the electrostatic energy difference, $E_{el} = 8E_C(n_{CPB} + n_{NR} - n - 1/2)$, between the nth and $(n+1)$th charge states, with the charging energy, $E_C = e^2/2C_\Sigma$, determined by the electron charge, e, and the CPB island's total capacitance, $C_\Sigma = C_{NR} + C_{CPB} + 2C_J$, where C_J is the capacitance of each Josephson junction and C_{CPB} is the capacitance between the CPB island and a nearby gate electrode. Here $n_{CPB} = C_{CPB}V_{CPB}/2e$ and $n_{NR} = C_{NR}V_{NR}/2e$ are the polarization charges (in units of Cooper pairs) applied by the gate electrode and the nanoresonator, which are held at potentials V_{CPB} and V_{NR}, respectively (Fig. 1b). The second term in $\hat{H}_{CPB}$ is the Josephson energy of the junctions, $E_J = E_{J0}|\cos(\pi\Phi/\Phi_o)|$, where Φ is the externally applied magnetic flux, $\Phi_o = h/2e$ is the flux quantum and E_{J0} is the maximum Josephson energy. From the diagonalization of $\hat{H}_{CPB}$ (ref. 21), we find the CPB ground, $|-\rangle$, and excited, $|+\rangle$, states to be separated by the transition energy $\Delta E = \sqrt{E_{el}^2 + E_J^2}$, where E_C/h and E_{J0}/h typically are ∼10 GHz.

Displacement (by x) of the nanoresonator results in linear modulation of the capacitance between the nanoresonator and CPB, $C_{NR}(x) \approx C_{NR}(0) + (\partial C_{NR}/\partial x)x$, which modulates the electrostatic energy of the CPB through n_{NR} and E_C, resulting in the interaction Hamiltonian[2] $\hat{H}_{int} = \hbar\lambda(\hat{a} + \hat{a}^\dagger)\hat{\sigma}_z$, where

$$\lambda \approx \frac{4n_{NR}E_C}{\hbar}\frac{1}{C_{NR}}\frac{\partial C_{NR}}{\partial x}x_{zp} \qquad (1)$$

is the capacitive coupling constant and $x_{zp} = \sqrt{\hbar/2M\omega_{NR}}$. For the parameter values used in this work (Supplementary Information), equation (1) yields $|\lambda/2\pi| \approx 0.3$–2.3 MHz.

The formal connection to CQED becomes clear when the full system Hamiltonian, $\hat{H} = \hat{H}_{NR} + \hat{H}_{CPB} + \hat{H}_{int}$, is transformed to the energy eigenbasis of the qubit:

$$\hat{H} = \hbar\omega_{NR}\hat{a}^\dagger\hat{a} + \frac{\Delta E}{2}\hat{\sigma}_z + \hbar\lambda(\hat{a} + \hat{a}^\dagger)\left(\frac{E_{el}}{\Delta E}\hat{\sigma}_z - \frac{E_J}{\Delta E}\hat{\sigma}_x\right) \qquad (2)$$

(where $\hat{\sigma}_z$ and $\hat{\sigma}_x$ are now Pauli matrices in the CPB's energy basis). Equation (2) is similar to a Jaynes–Cummings-type Hamiltonian[14]. With the qubit and nanoresonator far-detuned (that is, for
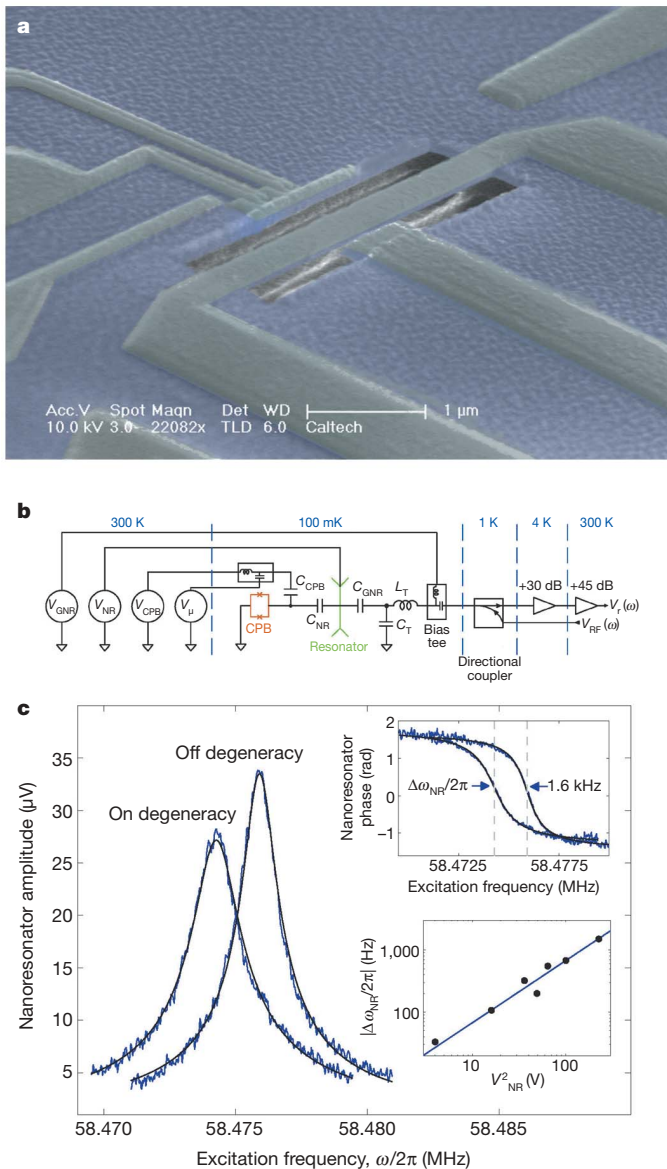
Figure 1 | Device and measurement circuit description, and driven frequency response of the nanoresonator. a, Colourized scanning electron micrograph of a device similar to the one measured. The nanoresonator is formed from low-stress silicon nitride with a thin coating (~80 nm) of aluminium for applying V_{NR}. The CPB is formed from aluminium during the same deposition steps as the nanoresonator. It is positioned at a distance ~300 nm from the nanoresonator, yielding the mutual capacitance $C_{NR} = 43$ aF. Adjacent to the CPB is an aluminium electrode for applying V_{CPB}. Another aluminium electrode is situated ~100 nm from the opposite side of the nanoresonator, for actuating the nanoresonator and measuring $\Delta\omega_{NR}/2\pi$. **b,** Circuit schematic for measuring $\Delta\omega_{NR}/2\pi$ using radio-frequency reflectometry (Supplementary Information). For typical values of the d.c. voltages V_{NR} and V_{GNR}, where V_{GNR} is applied to the actuation electrode and used to tune the coupling of the nanoresonator to the measurement circuit, the excitation signal, $V_{RF}(\omega)$, drives the nanoresonator at resonance ($\omega = \omega_{NR}$) to 1–10-pm root-mean-squared amplitude or an effective occupation of ~10^3–10^5 quanta. The nanoresonator's response is transformed by L_T and C_T for matching to a cryogenic amplifier. After amplification at room temperature (~300 K), the signal, $V_r(\omega)$, is fed to a radio-frequency lock-in for detection (Supplementary Information). **c,** The nanoresonator's amplitude (main panel) and phase (upper inset) versus excitation frequency, ω, for n_{CPB} biased on and off a charge degeneracy and $E_J/h \approx 10$ GHz. The solid black lines each denote a fit to a harmonic oscillator response. Lower inset: magnitude of the nanoresonator frequency shift, $|\Delta\omega_{NR}/2\pi|$ (black circles) as a function of V_{NR}^2 for $E_J/h \approx 11$–12 GHz and V_{CPB} biased at a charge degeneracy. The solid blue line is a fit to $|\Delta\omega_{NR}/2\pi| = AV_{NR}^2$, where A is a proportionality constant.

$\hbar|\lambda|\sqrt{N} \ll |\Delta E - \hbar\omega_{NR}|$), the dispersive coupling limit is realized, and, to lowest order, the system undergoes a shift in energy that can be viewed as a CPB-dressed correction to the nanoresonator's frequency[2]:

$$\frac{\Delta\omega_{NR}}{2\pi} = \frac{\hbar\lambda^2}{\pi} \frac{E_J^2}{\Delta E(\Delta E^2 - (\hbar\omega_{NR})^2)} \langle \hat{\sigma}_z \rangle \qquad (3)$$

For $\Delta E > \hbar\omega_{NR}$, $\Delta\omega_{NR}/2\pi < 0$ when the CPB resides in the ground state ($\langle \hat{\sigma}_z \rangle = -1$) and $\Delta\omega_{NR}/2\pi > 0$ when the CPB fully occupies the excited state ($\langle \hat{\sigma}_z \rangle = 1$). The dependence of $\Delta\omega_{NR}/2\pi$ on $\langle \hat{\sigma}_z \rangle$ is in close analogy to the single-atom refractive shift[14] that arises in the dispersive limit of CQED. In our system, $\Delta E \gg \hbar\omega_{NR}$, and it is appropriate to think of $\Delta\omega_{NR}/2\pi$ as arising solely from the CPB's state-dependent polarizability or 'quantum capacitance'[22,23]. Thus, for fixed E_J, $|\Delta\omega_{NR}/2\pi|$ is always maximized at CPB charge degeneracy points, $E_{el} = 0$, where the magnitude of the state-dependent component of the quantum capacitance is greatest.

We cool the sample to a temperature in the range of $T_{mc} \approx 100$–140 mK, where the qubit predominantly resides in the ground state (that is, $k_B T_{mc} \ll \Delta E$, where k_B is Boltzmann's constant) and the rate of quasiparticle poisoning in the qubit is minimal[24]. We then measure the nanoresonator frequency response using a combination of capacitive displacement transduction and radio-frequency reflectometry[25] (Fig. 1b and Supplementary Information). Figure 1c shows the frequency response of the nanoresonator amplitude (main panel) and phase (upper inset) at two values of V_{CPB} for fixed Φ and $V_{NR} = 15$ V (the largest coupling voltage used in the experiment). Consistent with equation (3) and the CPB residing in the ground state, when V_{CPB} is adjusted to a charge degeneracy point, the nanoresonator experiences a decrease in frequency, the magnitude of which is found to be $|\Delta\omega_{NR}/2\pi| \approx \hbar\lambda^2/\pi E_J = 1,600$ Hz. For fixed values of E_J and E_{el}, in agreement with equations (1) and (3), $|\Delta\omega_{NR}/2\pi|$ is found to exhibit a quadratic dependence on V_{NR} (Fig. 1c, lower inset) over the full range of V_{NR} values used in the experiment.

Embedding the nanoresonator in a phase-locked loop, we can track $\Delta\omega_{NR}/2\pi$ while keeping V_{NR} fixed and adiabatically sweeping V_{CPB} and Φ (Fig. 2a). The overall dependence of $\Delta\omega_{NR}/2\pi$ on V_{CPB} and Φ is in excellent qualitative agreement with our model (equation (3) and Fig. 2b–d). We find that $\Delta\omega_{NR}/2\pi$ exhibits the expected period-2e dependence on V_{CPB}, confirmed for four periods (Supplementary Information). We also observe that the periodicity of $\Delta\omega_{NR}/2\pi$ in Φ is in good agreement with one flux quantum Φ_o (Supplementary Information), as expected from the Φ dependence of E_J. At values of Φ for which $E_J/k_B \lesssim T_{mc}$ (for example trace 1 in Fig. 2c), the CPB excited state becomes thermally populated in the vicinity of the charge degeneracy points. As a result, the modulation depth of $\Delta\omega_{NR}/2\pi$ is reduced, which can be accounted for by replacing the qubit expectation in equation (3) with the Boltzmann-weighted average, $\langle \hat{\sigma}_z \rangle = -\tanh(\Delta E/2k_B T_{mc})$.

We can also manipulate the CPB state $\langle \hat{\sigma}_z \rangle$ by irradiating the CPB gate with microwaves that are resonant with the qubit transition, ΔE, and perform spectroscopy by monitoring the mechanical frequency shift, $\Delta\omega_{NR}/2\pi$. With the microwave frequency, $\omega_\mu/2\pi$, held fixed and the microwave amplitude, V_μ, adjusted such that polarization charge due to the microwave signal (in units of $2e$) satisfies $n_\mu = C_{CPB}V_\mu/2e \ll 1$, the CPB will oscillate between $|+\rangle$ and $|-\rangle$ with Rabi frequency $\Omega_d \approx 4E_C E_J n_\mu/\hbar\Delta E$ when V_{CPB} and Φ are tuned such that $\Delta E \approx \hbar\omega_\mu$. Because the response time of the nanoresonator, $2\pi/\kappa$, is long in comparison with characteristic timescales of the CPB's dynamics, measurements of $\Delta\omega_{NR}/2\pi$ will reflect the average qubit occupation, $\langle \hat{\sigma}_z \rangle = \rho_+ - \rho_-$, where ρ_+ and ρ_- are found from the steady-state solution to the Bloch equations[26]

$$\rho_+ = 1 - \rho_- = \frac{1}{2} \frac{\Omega_d^2 T_1 T_2}{1 + \Omega_d^2 T_1 T_2 + (\Delta E/\hbar - \omega_\mu)^2 T_2^2} \qquad (4)$$

and T_1 and T_2 are the qubit relaxation and dephasing times, respectively. For values of n_μ large enough that $\Omega_d^2 T_1 T_2 \gg 1$, the CPB
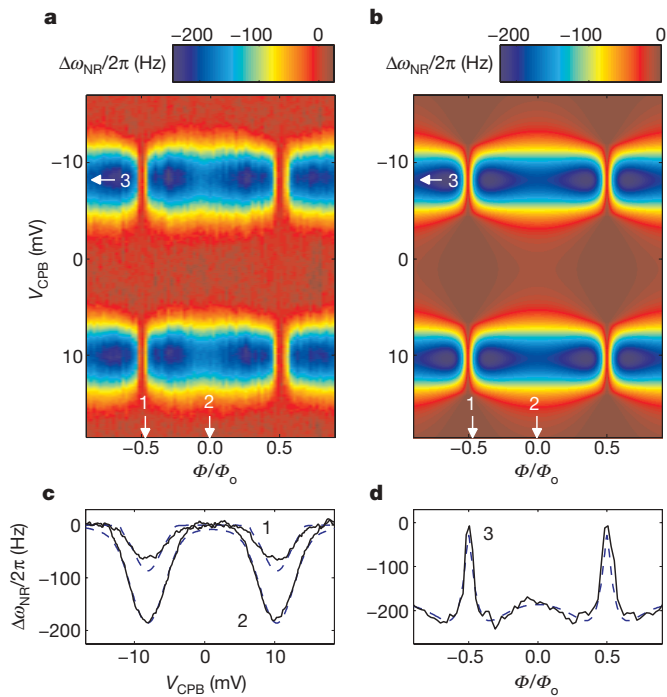
a

b

c

d

Figure 2 | Nanoresonator frequency shift as function of CPB parameters V_{CPB} and Φ/Φ_o. a, Measured $\Delta\omega_{NR}/2\pi$ for $V_{NR} = 7$ V and $T_{mc} \approx 100$ mK. Data has been post-processed to correct for charge drift and background fluctuations in $\omega_{NR}/2\pi$ (Supplementary Information). Normalization of the x axis is also discussed in the Supplementary Information. **b**, Numerically calculated $\Delta\omega_{NR}/2\pi$ as a function of V_{CPB} and Φ for $E_C/h = 14.0$ GHz, $E_{J0}/h = 13.2$ GHz and $|\lambda/2\pi| = 1.40$ MHz. The numerical model uses the full CPB Hamiltonian (Supplementary Information) to calculate the two lowest CPB eigenstates, $|+\rangle$ and $|-\rangle$. The CPB population, $\langle\hat{\sigma}_z\rangle$, is then calculated assuming the appropriate Boltzmann weighting. To account for low-frequency charge noise, $\Delta\omega_{NR}/2\pi$ from the model is convolved with a Gaussian of width $\sigma(2e) = 0.10$ in n_{CPB}. **c**, Comparison between data (solid black lines) and model (dashed blue lines) of selected traces of $\Delta\omega_{NR}/2\pi$ versus V_{CPB} for Φ biased near minimum E_J (labelled '1') and maximum E_J (labelled '2') . **d**, Comparison between data (solid black lines) and model (dashed blue lines) of $\Delta\omega_{NR}/2\pi$ versus Φ for V_{CPB} biased on a charge degeneracy (labelled '3').

becomes saturated, that is, $\rho_+ = \rho_- = 1/2$, and $\Delta\omega_{NR}/2\pi \rightarrow 0$. Thus, we can perform spectroscopy of the CPB by fixing $\omega_\mu/2\pi$ and n_μ and monitoring the nanomechanical frequency shift $\Delta\omega_{NR}/2\pi$ while adiabatically sweeping V_{CPB} and Φ (Fig. 3a–d). For $\omega_\mu/2\pi = 10.5$–20 GHz, we observe hyperbolae where $\Delta\omega_{NR}/2\pi \rightarrow 0$. These trace out constant-energy contours that are in general agreement with the expected n_{CPB}–Φ dependence of the qubit transition, ΔE (Fig. 3e). This allows us to extract the values $E_C/h = 12.7$–13.7 GHz and $E_{J0}/h \approx 13$ GHz (Supplementary Information), which, through equation (3), can be used to estimate the coupling strength, $|\lambda/2\pi| \approx 0.5$–3 MHz over the range $V_{NR} = 2$–15 V. Measurements of the qubit's linewidth, $\gamma/2\pi$, for varying microwave amplitude allow us to determine that $T_2 \gtrsim 2$ ns at charge degeneracy (Supplementary Information).

At large microwave amplitude V_μ ($n_\mu \gtrsim \pi E_J^2/16\hbar\omega_\mu E_C$), we demonstrate that we can utilize the nanomechanical frequency shift, $\Delta\omega_{NR}/2\pi$, as a probe of quantum coherent interference effects in the CPB (Fig. 4). These effects arise as a result of Landau–Zener tunnelling[27] that can occur between $|-\rangle$ and $|+\rangle$ whenever the CPB is swept, by means of V_μ, through the avoided-level crossing at charge degeneracy. If T_2 is greater than the microwave modulation period, $2\pi/\omega_\mu$, then successive Landau–Zener events can interfere, resulting in oscillations in the qubit population, $\langle\hat{\sigma}_z\rangle$, as a function of V_μ and V_{CPB}.

By monitoring $\Delta\omega_{NR}/2\pi$ while sweeping V_{CPB} at fixed values of V_μ, we clearly observe quantum interference (Fig. 4a). At the lowest

a

b

c

d

e
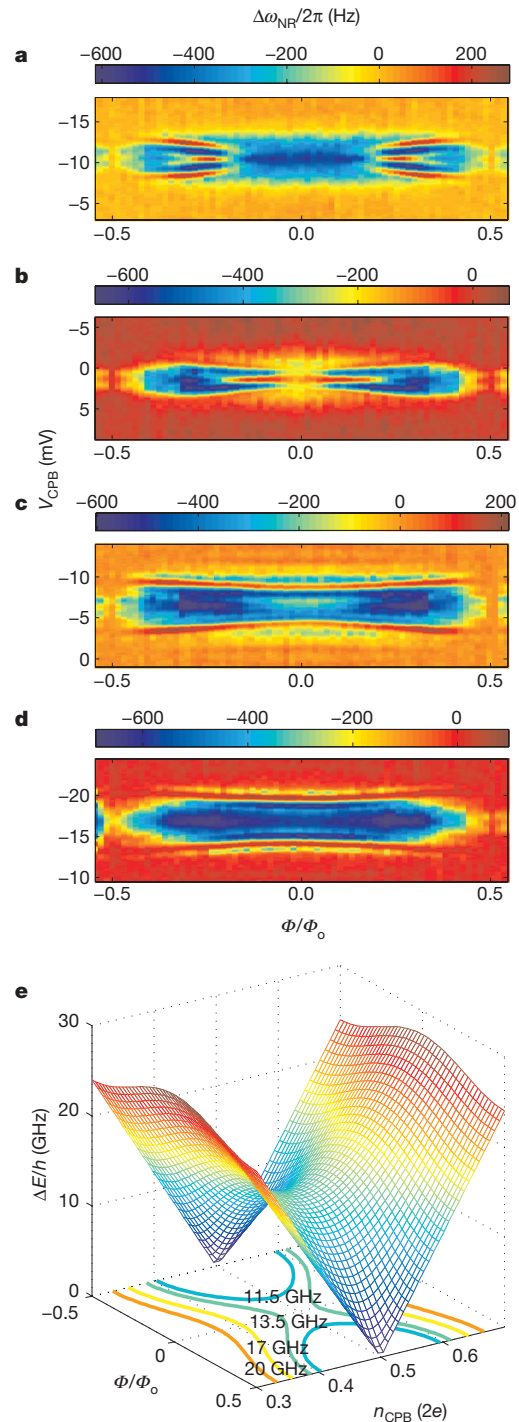
Figure 3 | Spectroscopy of the CPB using the nanomechanical frequency shift as a probe. a–d, $\Delta\omega_{NR}/2\pi$ measured as a function of V_{CPB} and Φ while applying microwaves of frequency $\omega_\mu/2\pi = 11.5$ GHz (**a**), 13.5 GHz (**b**), 17 GHz (**c**) and 20 GHz (**d**). Data has been post-processed to correct for charge drift and background fluctuations in $\omega_{NR}/2\pi$ (Supplementary Information). Normalization of the x axes is also discussed in the Supplementary Information. Data was taken for $V_{NR} = 10$ V and $T_{mc} \approx 140$ mK. **e**, Surface plot of CPB ground-state/excited-state splitting transition frequency, $\Delta E/h$, as a function of V_{CPB} and Φ, with constant energy contours at the microwave frequencies highlighted.

values of V_μ, Landau–Zener tunnelling is exponentially suppressed[27], and we observe a dependence of $\Delta\omega_{NR}/2\pi$ on V_{CPB} consistent with the CPB residing in $|-\rangle$. As V_μ is increased, we observe that $\Delta\omega_{NR}/2\pi$ oscillates with V_μ and V_{CPB}, even changing sign, and becoming maximally positive at values of V_μ and V_{CPB} for which we expect the occupation of $|+\rangle$ to be a maximum (the intersections of the
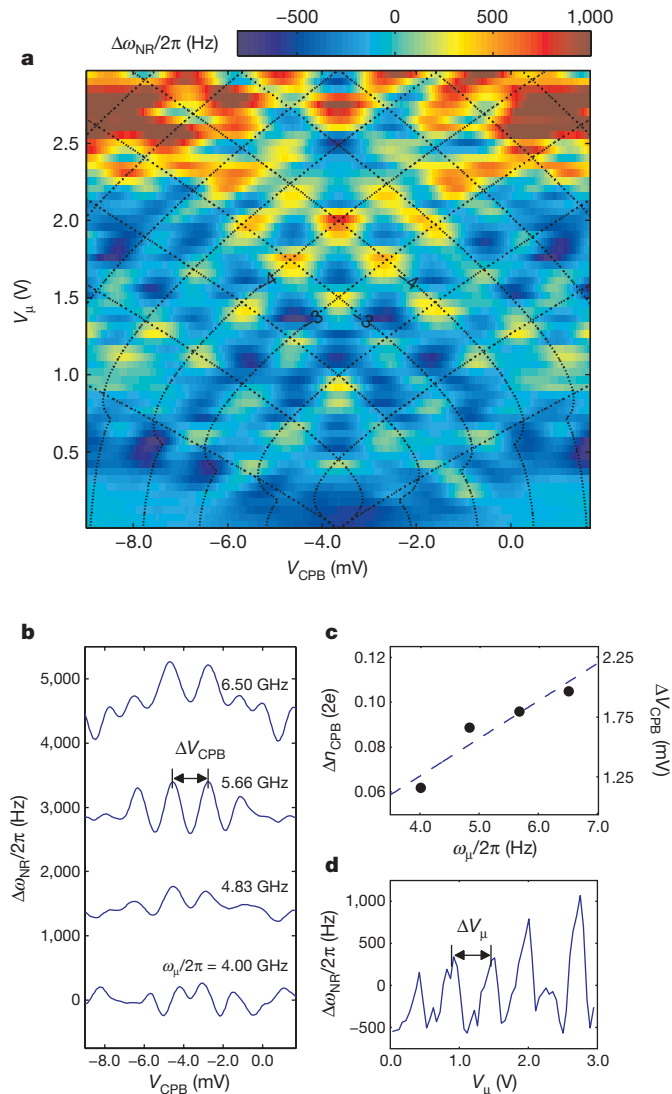
Figure 4 | Landau–Zener interferometry using the nanomechanical frequency shift as a probe. **a**, Interference fringes in $\Delta\omega_{NR}/2\pi$ plotted as a function of microwave amplitude, V_μ, and CPB gate voltage, V_{CPB}, for $\omega_\mu/2\pi = 6.50$ GHz. Data has been post-processed for charge drift (Supplementary Information). The colour scale is saturated at $\Delta\omega_{NR}/2\pi = +1,000$ Hz to enhance contrast of fringes at smaller values of V_μ. **b**, Cross-sections for constant values of V_μ, for $\omega_\mu/2\pi = 4.00$–6.50 GHz, chosen to coincide with the intersection of the $m = -3$ and $m = -4$ constant-phase contours, for $2\pi m$ advancement in the phase of the CPB wavefunction (Supplementary Information). The traces at different values of $\omega_\mu/2\pi$ have been offset vertically for clarity, and charge drift between data sets has been subtracted. **c**, Linear fit through the origin of the spacing, Δn_{CPB}, between adjacent interference fringes at the intersection of the $m = -3$ and $m = -4$ constant-phase contours. ΔV_{CPB} is determined from a fit of the interference fringes to a series of Gaussian peaks, and then converted to Δn_{CPB} using the CPB gate capacitance, $C_{CPB} = 17.1$ aF. Error bars are calculated from the Gaussian fit but are smaller than the point size and the scatter in the data, which is probably due to low-frequency charge noise. **d**, Nanomechanical frequency shift versus microwave amplitude for $\omega_\mu/2\pi = 6.50$ GHz at $V_{CPB} = -3.66$ mV, demonstrating the expected periodic modulation of the interference fringes. Data was taken for $V_{NR} = 10$ V and $T_{mc} \approx 110$ mK.

contours in Fig. 4a; Supplementary Information). We observe that the spacing, ΔV_{CPB}, in gate voltage between adjacent interference fringes increases linearly with increasing microwave frequency, $\omega_\mu/2\pi$, as expected[27] (Fig. 4b, c). A linear fit of Δn_{CPB} to $\omega_\mu/2\pi$ (Fig. 4c) yields $E_C/h = 14.9 \pm 0.6$ GHz (s.e.m.) in good agreement with the value extracted from spectroscopy. Figure 4d shows a cross-section of $\Delta\omega_{NR}/2\pi$ as a function of V_μ at charge degeneracy, demonstrating

the expected periodic dependence of the interference maxima. The primary maxima in $\Delta\omega_{NR}/2\pi$ occur for values of V_μ that produce a phase shift of $2\pi m$ (where m is an integer) in the CPB's wavefunction over one-half cycle of microwave modulation. The resulting average spacing between peaks, ΔV_μ, found from a fit of the data to a series of Lorentzians, provides an estimated total attenuation of 45 ± 2 dB (s.e.m.) at $\omega_\mu/2\pi = 6.50$ GHz in the CPB gate line, which is in reasonable agreement with measurements of the attenuation made before cool-down with the apparatus at ~ 300 K (~ 50–54 dB). It should be possible to extract the qubit dephasing time, T_2, from the width of the interference fringes by using a model that carefully considers the various timescales in the problem (that is, T_1, $2\pi/\omega_\mu$ and $2\pi/\omega_{NR}$)[27].

For both driven and non-driven CPB cases, it is notable how well the simple dispersive model (equation (2)) agrees with our observations. It is not obvious, a priori, that the equations of motion used to model the interaction between an atom and a photon should also apply to the interaction between a suspended nanostructure and a mesoscopic electronic device, in particular because the latter systems each comprise billions of atoms. Despite this agreement, several outstanding issues are noteworthy. First, we observe increased damping of the NEMS upon tuning the CPB to the charge degeneracy point. Although further explorations are necessary to determine the origin of this excess energy loss, the fact that it depends on the CPB gate bias, V_{CPB}, and increases with V_{NR} suggests that it is mediated by the CPB. Second, we observe additional resonant features near charge degeneracy (Fig. 3 and Supplementary Information) whose origins are not yet understood. These robust features do not appear to be sensitive to time or background electric field. Furthermore, they also do not demonstrate a clear dependence on n_μ, suggesting that mechanisms such as multiphoton transitions[19] and Landau-Zener tunnelling[27] may be ruled out.

The dispersive interaction that we have measured, in conjunction with techniques that have been used to manipulate[28] and measure[18,28] superconducting qubits, could soon be used to generate and probe entangled states of nanomechanical systems and qubits. For example, a superposition of nanoresonator coherent states oscillating at distinct frequencies dressed by the state of the CPB (that is, $\omega_{NR,\pm}/2\pi = \omega_{NR}/2\pi \pm \Delta\omega_{NR}/2\pi$) could be generated by dispersively coupling the nanoresonator to a CPB that is initially prepared in a superposition of $|-\rangle$ and $|+\rangle$ (refs 12, 13). Coherence of the nanomechanical system would manifest itself in periodic reductions in and revivals of the coherent oscillations of the qubit as the phases of the two nanoresonator states shift out of alignment and back. This could then be quantified either through careful measurements of the qubit's dephasing spectrum[12,18] or using qubit 'echo' techniques[13,28].

Theoretical investigations of the second approach suggest that entanglement 'recoherences' should be observable in systems similar to ours using a coupling strength of $|\lambda/2\pi| \approx 10$ MHz (ref. 13). This would require a modest improvement to the existing sample, which we anticipate is achievable by engineering a smaller gap electrode; for example, an order-of-magnitude increase in λ is expected using parameters similar to those already demonstrated with single-electron transistors[29]. It will also be necessary to reduce the effects of quasiparticle poisoning and other sources of charge noise to achieve CPB dephasing times $T_2 \gtrsim 100$ ns. This has been accomplished in circuit QED through careful engineering of the CPB's parameters and by using a superconducting cavity for isolation and measurement of the CPB[18,28]. Reducing quasiparticle poisoning will have the additional benefit of enabling operation of the experiment at lower temperatures, where the deleterious effects of the nanoresonator's thermal fluctuations on the visibility of qubit revivals should be much weaker. The calculations in ref. 13 indicate that recoherences should be observable for nanoresonator thermal occupation factors up to ~ 20, which corresponds to $T_{mc} \lesssim 60$ mK for $\omega_{NR}/2\pi \approx 60$ MHz. This is readily attainable with our dilution refrigerator.

We have demonstrated the read-out of a superconducting qubit using a dispersive interaction with a nanomechanical resonator. This technique joins cantilever-based magnetic resonance force detection[30] as the only demonstrated mechanical probe techniques of individual two-level quantum systems. The realistic prospects of investigating quantum coherence in a nanomechanical resonator establish the nanoresonator-coupled qubit as a valuable new tool with which to explore further the frontiers of quantum mechanics.

1. Armour, A. D., Blencowe, M. P. & Schwab, K. C. Entanglement and decoherence of a micromechanical resonator via coupling to a Cooper-pair box. *Phys. Rev. Lett.* **88**, 148301 (2002).
2. Irish, E. K. & Schwab, K. C. Quantum measurement of a coupled nanomechanical resonator-Cooper-pair box system. *Phys. Rev. B* **68**, 155311 (2003).
3. Martin, I., Shnirman, A., Tian, L. & Zoller, P. Ground state cooling of mechanical resonators. *Phys. Rev. B* **69**, 125339 (2004).
4. Cleland, A. N. & Geller, M. R. Superconducting qubit storage and entanglement with nanomechanical resonators. *Phys. Rev. Lett.* **93**, 070501 (2004).
5. Tian, L. Entanglement from a nanomechanical resonator weakly coupled to a single Cooper-pair box. *Phys. Rev. B* **72**, 195411 (2005).
6. Buks, E. & Blencowe, M. P. Decoherence and recoherence in a vibrating rf SQUID. *Phys. Rev. B* **74**, 174504 (2006).
7. Wei, L. F. Liu, Y.-x., Sun, C. P. & Nori, F. Probing tiny motions of nanomechanical resonators: classical or quantum mechanical? *Phys. Rev. Lett.* **97**, 237201 (2006).
8. Jacobs, K., Lougovski, P. & Blencowe, M. P. Continuous measurement of the energy eigenstates of a nanomechanical resonator without a non-demolition probe. *Phys. Rev. Lett.* **98**, 147201 (2007).
9. Clerk, A. A. & Utami, D. W. Using a qubit to measure photon-number statistics of a driven thermal oscillator. *Phys. Rev. A* **75**, 042302 (2007).
10. Hauss, J. *et al.* Single-qubit lasing and cooling at the Rabi frequency. *Phys. Rev. Lett.* **100**, 037003 (2008).
11. Jacobs, K., Jordan, A. N. & Irish, E. K. Energy measurements and preparation of canonical phase states of a nano-mechanical resonator. *Europhys. Lett.* **82**, 18003 (2008).
12. Utami, D. W. & Clerk, A. A. Entanglement dynamics in a dispersively coupled qubit oscillator system. *Phys. Rev. A* **78**, 042323 (2008).
13. Armour, A. D. & Blencowe, M. P. Probing the quantum coherence of a nanomechanical resonator using a superconducting qubit: I. Echo scheme. *N. J. Phys.* **10**, 095004 (2008).
14. Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
15. Heinzen, D. J. & Feld, M. S. Vacuum radiative level shift and spontaneous-emission linewidth of an atom in an optical resonator. *Phys. Rev. Lett.* **59**, 2623–2626 (1987).
16. Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
17. Brune, M. *et al.* Observing the progressive decoherence of the "meter" in a quantum measurement. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
18. Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
19. Wilson, C. M. *et al.* Coherence times of dressed states of a superconducting qubit under extreme driving. *Phys. Rev. Lett.* **98**, 257003 (2007).
20. Nakamura, Y., Pashkin, A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
21. Makhlin, Y., Schon, G. & Shnirman, A. Quantum-state engineering with Josephson junction devices. *Rev. Mod. Phys.* **73**, 357–400 (2001).
22. Sillanpaa, M. A. *et al.* Direct observation of Josephson capacitance. *Phys. Rev. Lett.* **95**, 206806 (2005).
23. Duty, T. *et al.* Observation of quantum capacitance in the Cooper-pair transistor. *Phys. Rev. Lett.* **95**, 206807 (2005).
24. Palmer, B. S. *et al.* Steady-state thermodynamics of non-equilibrium quasiparticles in a Cooper-pair box. *Phys. Rev. B* **76**, 054501 (2007).
25. Truitt, P. A. *et al.* Efficient and sensitive capacitive readout of nanomechanical resonator arrays. *Nano Lett.* **7**, 120–126 (2007).
26. Schuster, D. I. *et al.* ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
27. Sillanpaa, M. *et al.* Continuous-time monitoring of Landau-Zener interference in a Cooper-pair box. *Phys. Rev. Lett.* **96**, 187002 (2006).
28. Leek, P. J. *et al.* Observation of Berry's phase in a solid-state qubit. *Science* **318**, 1889–1892 (2007).
29. Naik, A. *et al.* Cooling a nanomechanical resonator with quantum back-action. *Nature* **443**, 193–196 (2006).
30. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004).

LETTERS

Peierls distortion as a route to high thermoelectric performance in In$_4$Se$_{3-\delta}$ crystals

Jong-Soo Rhyee[1], Kyu Hyoung Lee[1], Sang Mock Lee[1], Eunseog Cho[1], Sang Il Kim[1], Eunsung Lee[1], Yong Seung Kwon[2], Ji Hoon Shim[3] & Gabriel Kotliar[4]

Thermoelectric energy harvesting—the transformation of waste heat into useful electricity—is of great interest for energy sustainability. The main obstacle is the low thermoelectric efficiency of materials for converting heat to electricity, quantified by the thermoelectric figure of merit, *ZT*. The best available n-type materials for use in mid-temperature (500–900 K) thermoelectric generators have a relatively low *ZT* of 1 or less, and so there is much interest in finding avenues for increasing this figure of merit[1]. Here we report a binary crystalline n-type material, In$_4$Se$_{3-\delta}$, which achieves the *ZT* value of 1.48 at 705 K—very high for a bulk material. Using high-resolution transmission electron microscopy, electron diffraction, and first-principles calculations, we demonstrate that this material supports a charge density wave instability which is responsible for the large anisotropy observed in the electric and thermal transport. The high *ZT* value is the result of the high Seebeck coefficient and the low thermal conductivity in the plane of the charge density wave. Our results suggest a new direction in the search for high-performance thermoelectric materials, exploiting intrinsic nanostructural bulk properties induced by charge density waves.

Over the past decade, there has been a renewed focused effort on thermoelectric materials motivated by the increasing societal needs for renewable energy and by technological advances in nanoscale science. The thermoelectric performance of a given material is characterized by the materials' dimensionless figure of merit *ZT* ($ZT = S^2\sigma T/\kappa$, where *Z*, *S*, σ, *T*, and κ are respectively a measure of the material's thermoelectric properties, Seebeck coefficient, electrical conductivity, absolute temperature, and thermal conductivity), which can be enhanced by following two approaches. The first seeks to minimize the thermal conductivity by promoting phonon scattering and localization while preserving the itinerancy of the electron transport[2–6]. The second seeks to enhance the power factor $S^2\sigma$ through the quantum confinement effect in reduced dimensionalities[7,8].

Recently, a new approach to ultralow thermal conductivity was explored. It was suggested that the layered structure of disordered two-dimensional crystalline sheets may have extremely low thermal conductivity[9]. Based on this, we propose the possible application of charge density waves (CDWs) to thermoelectricity. The CDW is a low-dimensional transport phenomenon involving strong electron–phonon coupling[10]. This strong interaction breaks the translational symmetry of lattices. The in-plane lattice distortion by CDWs of layered structured materials may realize the concept of layered and disordered crystalline sheets. In spite of the possible low thermal conductivity in a CDW system, it is necessary to select one type of carrier transport (electron or hole) to achieve a high power factor, because mixed carrier transport (electrons and holes) suppresses the Seebeck coefficient[11]. In$_4$Se$_3$ provides an interesting realization of a material

with reduced dimensionality, supporting quasi-one-dimensional In-chains[12], nanorod structures[13], and asymmetric band dispersion[14]. The crystal structure of In$_4$Se$_3$ is shown in Fig. 1. It forms a layered structure of (In$_3$)$^{5+}$ clusters covalently bonded to Se ions in the *b*–*c* planes held together by van der Waals interactions along the *a* axis. Although the layered crystal structure is similar to the conventional Bi$_2$Te$_3$ compound, the intrinsic properties of the nanorod structure and the quasi-one-dimensional chains of In$_4$Se$_3$ are different from those of Bi$_2$Te$_3$ and similar alloyed compounds. It is a bandgap semiconductor having a sizable bandgap of 0.5–1.0 eV with anisotropic band dispersions[14]. To reduce the energy of the bandgap, we employed self doping by Se deficiency, that is, a compound of formula In$_4$Se$_{3-\delta}$. We successfully grew two Se-deficient In$_4$Se$_{3-\delta}$ crystal ingots; In$_{59}$Se$_{41}$ (In$_4$Se$_{2.78}$, $\delta = 0.22$) and In$_{63}$Se$_{37}$ (In$_4$Se$_{2.35}$, $\delta = 0.65$). From X-ray diffraction of the cross-sectional plane of the grown crystal, we find that the growth direction contains the *a*–*b* plane, whereas the perpendicular to the growth direction has two mixed crystalline orientations of *b*–*c* and *a*–*c* planes (Supplementary Information); here we denote the perpendicular to the growth direction as the *b*–*c* plane, for convenience, because electrical conduction is thought to be dominant along the *b*–*c* plane, based on the results of band structure calculation of In$_4$Se$_{3-\delta}$ ($\delta = 0.25$).

Figure 2 shows representative thermoelectric properties of In$_4$Se$_{2.35}$ ($\delta = 0.65$) along the growth direction (*a*–*b* plane, squares) and perpendicular to the growth direction (*b*–*c* plane, red circles), together with results from theoretical Boltzmann transport calculations. As shown in Fig. 2a, the thermal conductivity $\kappa(T)$ of In$_4$Se$_{2.35}$ is very low ($\leq$1.2 W m^{-1} K^{-1} at 300 K) along the *b*–*c* plane, and it
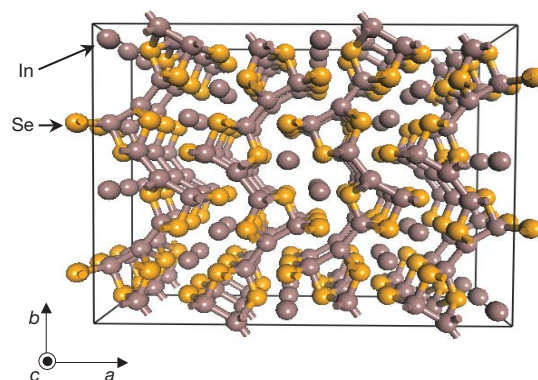

Figure 1 | Crystal structure of In$_4$Se$_3$. Perspective view of the *a*–*b* plane. Covalently bonded In–Se layers are stacked along the *a*-axis direction by relatively strong van der Waals interactions.
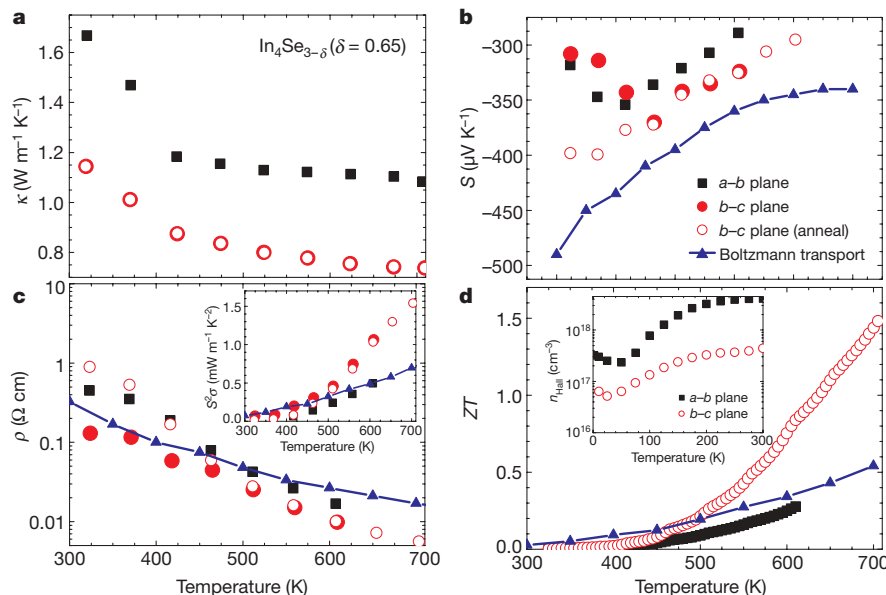
965

Figure 2 | Anisotropic thermoelectric properties, annealing effect, and angle-averaged value of Boltzmann transport calculations for In$_4$Se$_{3-\delta}$ ($\delta = 0.65$) bulk crystal. Shown are thermoelectric properties along the a–b (squares) and b–c planes (circles). Symbols shown both open and filled indicate measurements before and after heat treatment (450 °C, 24 h), respectively. Blue triangles and lines indicate the result of the Boltzmann transport calculations. **a**, Temperature-dependent thermal conductivities, $\kappa(T)$. **b**, Seebeck coefficient, $S(T)$. **c**, Electrical resistivity, $\rho(T)$, and power factor defined by $S^2\sigma$ (inset). **d**, Dimensionless figure-of-merit ZT and effective carrier concentration, n_{eff} (inset).

decreases with increasing temperature (0.74 W m^{-1} K^{-1} at 705 K). The thermal conductivities before and after heat treatment (450 °C for 24 h) are almost identical along the b–c plane (open red circles).

The Seebeck coefficient $S(T)$ and electrical resistivity $\rho(T)$ are presented in Fig. 2b and 2c, respectively. Compounds of formula In$_4$Se$_{3-\delta}$ ($\delta \leq 0.65$) are n-type materials with negative Seebeck coefficient. Before heat treatment, In$_4$Se$_{3-\delta}$ ($\delta = 0.65$) exhibits a sharp dip in S near 415 K (a–b plane) and 465 K (b–c plane) which is close to the bulk melting temperature of the excess indium ($T_m \approx 430$ K). The low temperature thermal instability can be eliminated by annealing the sample at 450 °C for 24 h, which gives reproducible data without the transitions, as shown in Fig. 2b (red open circles). The electrical resistivity $\rho(T)$ before and after heat treatment shows semiconducting behaviour of $\rho(T)$ (Fig. 2c). Although the low-temperature values of $S(T)$ and $\rho(T)$ are affected by the excess In, the annealing effect is not significant in the regime which is relevant for thermoelectricity at temperatures higher than T_m.

The dimensionless figure of merit ZT shown in Fig. 2d reaches the remarkably high value of 1.48 for In$_4$Se$_{2.35}$ ($\delta = 0.65$) at 705 K along the b–c plane. In the case of In$_4$Se$_{2.78}$ ($\delta = 0.22$), ZT reaches 1.1 at the same temperature (Supplementary Information). The higher ZT values of In$_4$Se$_{2.35}$ ($\delta = 0.65$) than those of In$_4$Se$_{2.78}$ ($\delta = 0.22$) are mainly caused by the lowering of thermal conductivity with increasing Se deficiency owing to the defect-induced phonon scattering at the Se-defect sites. The energy recycling efficiency η is calculated from the relation[1];

$$\eta = \frac{\Delta T}{T_h} \frac{\sqrt{1+ZT}-1}{\sqrt{1+ZT}+\frac{T_c}{T_h}}$$

where T_h (T_c) is the hot(cold)-side temperature and $\Delta T/T_h$ is the Carnot efficiency. The efficiency η for waste heat recovery of In$_4$Se$_{2.35}$ ($\delta = 0.65$) can reach 11.4% when T_h is 705 K and the temperature difference $\Delta T = 300$ K. It could generate commercially viable levels of power by recycling waste heat in this mid-temperature range. The Hall carrier concentration n_{Hall} of the In$_4$Se$_{2.35}$ ($\delta = 0.65$) crystal for crystal orientations along the a–b and the b–c planes, determined by Hall resistivity measurements, is estimated to be 4×10^{18} cm^{-3} and 4×10^{17} cm^{-3} at 300 K, respectively (Fig. 2d inset). The carrier concentrations for the In$_4$Se$_{3-\delta}$ compounds of $\delta = 0.22$ and $\delta = 0.65$

(7×10^{18} cm^{-3} along the a–b plane for $\delta = 0.22$) are lower than the optimum carrier concentration (10^{19}–10^{20} cm^{-3}) for narrow bandgap semiconductors[15].

To understand the microscopic origin of the exceptional properties of this material, we performed first-principles transport calculations of In$_4$Se$_3$ using the BoltzTraP program[16]. Calculations of the angle-averaged transport coefficients, within the rigid band approximation, can account for the experimental trends and confirm the potential of In$_4$Se$_{3-\delta}$ as a thermoelectric material. Using a chemical potential $\mu = 0.22$ eV and a scattering time $\tau = 2.2 \times 10^{-14}$ s, we obtained an electron concentration of 1×10^{18} cm^{-3} and a high Seebeck coefficient (-360 μV K^{-1} at 600 K), which is comparable to the observed experimental value. The calculated temperature-dependent behaviour of $S(T)$ and $\rho(T)$ can qualitatively reproduce the experimental data, as shown in Fig. 2b and c (blue line and blue triangles). Using a constant value of the angle-averaged lattice thermal conductivity $\kappa_{\text{ph}} = 0.8$ W m^{-1} K^{-1}, we can account for the temperature dependence of the power factor and ZT, as shown in the inset of Fig. 2c and Fig. 2d, respectively. Within the rigid band approximation, the carrier density of those compounds is not yet optimal (Supplementary Information), and therefore ZT could be further increased by optimizing the carrier concentration by means of chemical substitutions.

High-resolution transmission electron microscopy (HRTEM) images and electron diffraction patterns are presented in Fig. 3. Figure 3a shows the HRTEM image of the a–b plane of the In$_4$Se$_{2.78}$ ($\delta = 0.22$) crystal. The Bragg spots of the electron diffraction pattern in the a–b plane (Fig. 3a inset) accompany small superstructure peaks in the chain direction. Quasi-one-dimensional Bragg spots and secondary superstructure peaks indicate the lattice distortion along the chain direction. Figure 3b shows the HRTEM image of the cross-sectional plane of the In$_4$Se$_{2.78}$ ($\delta = 0.22$) crystal ingot. There are several grain boundaries between the stripe and chequerboard patterns with arbitrary orientation: the stripe and chequerboard patterns are considered to be the a–c and b–c planes, respectively. The grains are about 20 nm or less in width. The electron diffraction pattern of the b–c plane shows a rhomboidal Bragg diffraction pattern with one-dimensional superlattice peaks between the bright Bragg peaks (Fig. 3b inset), indicating the presence of a Peierls instability along the b-axis direction.
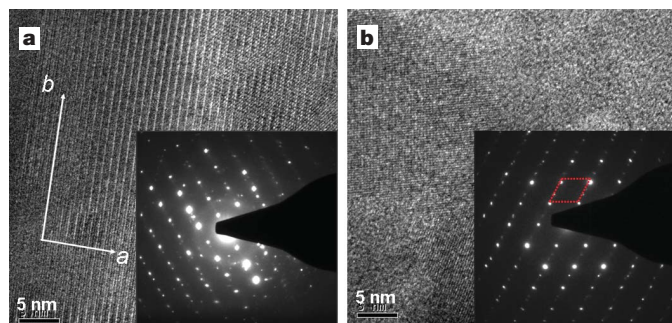
Figure 3 | High-resolution TEM images and electron diffraction patterns of In$_4$Se$_{3-\delta}$ ($\delta = 0.22$). **a**, HRTEM image of the a–b plane. The one-dimensional chains are periodically aligned along the b direction. Inset, the electron diffraction pattern in the a–b plane. The diffraction patterns of the quasi-one-dimensional chains are revealed along the b direction. Bragg spots with superstructure peaks along the b direction are due to the commensurate lattice distortions. **b**, HRTEM image of the cross-sectional plane of the grown crystal. Arbitrarily aligned chequerboard and stripe phases are of b–c and a–c planes. Inset, electron diffraction pattern of the b–c plane. Red rhomboid indicates the diffraction pattern of unit cell.

The Peierls instability of this material is also suggested by theoretical considerations, once we explicitly include the Se vacancies in the calculations. We calculated the generalized electron susceptibility $\chi(q)$, where q is the momentum vector (ref. 17), from the band structure of In$_4$Se$_{3-\delta}$ ($\delta = 0.25$); $\chi(q)$ is shown along the X–U symmetry line in Fig. 4b. The sharp peak at the (0, 1/2, 1/16) point can be understood from the quasi-one-dimensional Fermi surface nesting in the b–c plane shown in Fig. 4a. It consists of two smooth diamond-shaped Fermi surfaces located at the upper and lower zone boundaries of the first Brillouin zone. There is a well defined commensurate nesting vector **q** (red arrow) defined in the b–c plane which can result in a CDW instability once the electron–phonon or electron–electron correlations are incorporated in a calculation going beyond the local-density approximation. Because of the long range lattice modulation along the c direction, the density wave instability is closely connected to the quasi-one-dimensional Peierls instability of the chain-like structure along the b axis of this material.

The electron diffraction pattern shown in Fig. 3b inset is consistent with this Fermi surface nesting behaviour. The small faint peaks between the Bragg spots indicate the doubling superstructure in the b direction, while the long-range modulation in the c direction cannot be seen, probably owing to the experimental resolution. Lattice distortion along the b axis is clearly visible in the electron diffraction pattern shown in Fig. 3a inset.

Thermal transport properties shown in Fig. 2 also suggest the presence of a CDW. We note that the thermal conductivity in the b–c plane is lower than that in the a–b plane in the In$_4$Se$_{2.35}$ crystal, which is at first sight surprising, as the bonding in the b–c plane is weaker than that in the a–b plane. The in-plane (b–c plane) lattice distortion driven by the CDW lowers the thermal conductivity. In addition, the Hall carrier concentration is anisotropic with respect to crystal orientations. In usual cases, the carrier density is isotropic with respect to crystal orientations. However, in this case, the carrier density along the b–c plane is lower than that of the a–b plane, which is caused by the CDW along the b–c plane. Because of the CDW gap opening, the itinerant carrier density is decreased along the b–c plane. These facts, and the $S(T)$ and $\rho(T)$ shown in Fig. 2b and c, suggest the formation of a CDW at a temperature higher than 705 K.

The theoretical calculations in the rigid band approximation of the In$_4$Se$_{3-\delta}$ compound cannot however account for the observed anisotropy of the transport coefficients (see Supplementary Information for details). The rigid band approximation does not take into account the CDW with supercell zone boundaries. The quasi-one-dimensional lattice distortion, and the exotic anisotropic properties of thermal and electrical transport coefficients, together indicate the important role of the CDW in achieving a high ZT in the b–c plane. The in-plane lattice distortion by CDW lowers thermal conductivity, and the intrinsic nanostructural low dimensionality of this material results in the high Seebeck coefficient.

The present thermoelectric investigations of In$_4$Se$_{3-\delta}$ crystals suggest that bulk low-dimensional layered materials with strong electron–phonon coupling, such as Peierls or CDW instabilities, are promising candidates for new thermoelectric materials. Compared to other realizations of high-ZT n-type thermoelectric materials through nanoscale phase separation[18,19], the CDW mechanism for high ZT has the advantage of being realized as an intrinsically bulk phenomenon without artificial control of nanoscale phase segregation. This resulted in an exceptionally high ZT (1.48 at 705 K) in compounds with high chemical stability and good mechanical properties that can be combined with well established p-type thermoelectrics (such as skutterudites) to produce highly efficient thermoelectric power generation modules.

METHODS SUMMARY

The In$_4$Se$_{3-\delta}$ crystal ingots were grown by the Bridgeman method. In and Se were placed in an evacuated quartz ampoule with an excess of In (5–10 at.%) for Se-deficiency control. Heat treatment was followed by the melting and crystallizing at 550 °C and 590 °C for $\delta = 0.65$ and 0.22 compounds, respectively, during a week with a growth rate of 1.5 mm h^{-1}. The chemical inhomogeneity was examined by inductively coupled plasma spectroscopy and electron dispersive spectroscopy measurements. High-temperature thermal conductivity was obtained by measurements of sample density ρ_s, thermal diffusivity λ (by the laser flash method), and heat capacity C_p (in an ULVAC system); $\kappa = \rho\lambda C_p$, where heat capacity C_p was derived from Dulong-Petit fitting at high temperatures ($T \geq 300$ K). The high-temperature electrical resistivity and Seebeck coefficient were measured by the four-probe method (in an ULVAC system). The Hall resistivity ρ_{xy} measurement was carried out by the five-contact AC-transport technique, using a physical property measurement system (Quantum Design). The effective carrier concentration was calculated by the one-band model as the following relation: $n_{eff} = -1/(R_H e)$, where the Hall coefficient $R_H = \rho_{xy}/H$ and $e = 1.602 \times 10^{-19}$ C.

The first-principles calculation was performed by the pseudopotential plane wave method using the Vienna *Ab initio* Simulation Package. We adopted the generalized gradient approximation implemented by Perdew, Burke, and Ernzerhof for the exchange correlation energy functional with the spin–orbit interaction. The thermoelectric properties are calculated using the BoltzTraP program.

Figure 4 | Fermi surface and generalized electron susceptibility of In$_4$Se$_{3-\delta}$. **a**, Fermi surface of In$_4$Se$_{3-\delta}$ ($\delta = 0.25$) in the b–c plane (blue and green lines). Black square is the first Brillouin zone. Fermi nesting vector **q** (red arrow) is defined in the closed Fermi surface. **b**, Generalized electron susceptibility $\chi(q)$ along the X(0, 1/2, 0)–U(0, 1/2, 1/2) symmetry line (top). The $\chi(q)$ has a singularity at the (0, 1/2, 1/16) point (bottom). The numbers on the horizontal axis indicates the calculation mesh along the X–U–X symmetry line.

1. Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nature Mater.* **7**, 105–114 (2008).
2. Sales, B. C. Electron crystals and phonon glasses: a new path to improved thermoelectric materials. *Mater. Res. Soc. Bull* **23**, 15–21 (1998).

3. Snyder, G. J., Christensen, M., Nishibori, E. J., Caillat, T. & Iversen, B. B. Disordered zinc in Zn_4Sb_3 with phonon-glass and electron-crystal thermoelectric properties. *Nature Mater.* **3**, 458–463 (2004).
4. Wölfing, B., Kloc, C., Teubner, J. & Bucher, E. High performance thermoelectric Tl_9BiTe_6 with an extremely low thermal conductivity. *Phys. Rev. Lett.* **86**, 4350–4353 (2001).
5. Boukai, A. I. *et al.* Silicon nanowires as efficient thermoelectric materials. *Nature* **451**, 168–171 (2007).
6. Venkatasubramanian, R., Siivola, E., Colpitts, T. & O'Quinn, B. Thin-film thermoelectric devices with high room-temperature figures of merit. *Nature* **413**, 597–602 (2001).
7. Dresselhaus, M. S. *et al.* New directions for low-dimensional thermoelectric materials. *Adv. Mater.* **19**, 1043–1053 (2007).
8. Lin, Y.-M. & Dresselhaus, M. S. Thermoelectric properties of superlattice nanowires. *Phys. Rev. B* **68**, 075304 (2003).
9. Chiritescu, C. *et al.* Ultralow thermal conductivity in disordered, layered WSe_2 crystals. *Science* **315**, 351–353 (2007).
10. Güner, G. *Density Waves in Solids* (Addison-Wesley, 1994).
11. Rhyee, J. S. *et al.* Thermal and electronic transport properties of $CeTe_{2-x}Sn_x$ compounds. *J. Appl. Phys.* **105**, 053712 (2009).
12. Losovyj, Y. B. *et al.* The electronic structure of surface chains in the layered semiconductor In_4Se_3 (100). *Appl. Phys. Lett.* **92**, 122107 (2008).
13. Balitskii, O. A., Savchyn, V. P., Jaeckel, B. & Jaegermann, W. Surface characterization of In_4Se_3 single crystals. *Physica E* **22**, 921–923 (2004).
14. Losovyj, Y. B. *et al.* The anisotropic band structure of layered In_4Se_3 (001). *J. Appl. Phys.* **104**, 083713 (2008).
15. Mahan, G., Sales, B. & Sharp, J. Thermoelectric materials: new approaches to an old problem. *Phys. Today* **50**, 42–47 (March 1997).
16. Madsen, G. K. H. & Singh, D. J. BoltzTraP. A code for calculating band-structure dependent quantities. *Comput. Phys. Commun.* **175**, 67–71 (2006).
17. Shim, J. H., Kang, J.-S. & Min, B. I. Electronic structures of RTe_2 (R=La, Ce): a clue to the pressure-induced superconductivity in $CeTe_{1.82}$. *Phys. Rev. Lett.* **93**, 156406 (2004).
18. Hsu, K. F. *et al.* Cubic $AgPb_mSbTe_{2+m}$: bulk thermoelectric materials with high figure of merit. *Science* **303**, 818–821 (2004).
19. Chen, N. *et al.* Macroscopic thermoelectric inhomogeneities in $(AgSbTe_2)_x(PbTe)_{1-x}$. *Appl. Phys. Lett.* **87**, 171903 (2005).

Increased seasonality through the Eocene to Oligocene transition in northern high latitudes

James S. Eldrett[1], David R. Greenwood[2], Ian C. Harding[3] & Matthew Huber[4]

A profound global climate shift took place at the Eocene–Oligocene transition (~33.5 million years ago) when Cretaceous/early Palaeogene greenhouse conditions gave way to icehouse conditions[1–3]. During this interval, changes in the Earth's orbit and a long-term drop in atmospheric carbon dioxide concentrations[4–6] resulted in both the growth of Antarctic ice sheets to approximately their modern size[2,3] and the appearance of Northern Hemisphere glacial ice[7,8]. However, palaeoclimatic studies of this interval are contradictory: although some analyses indicate no major climatic changes[9,10], others imply cooler temperatures[11], increased seasonality[12,13] and/or aridity[12–15]. Climatic conditions in high northern latitudes over this interval are particularly poorly known. Here we present northern high-latitude terrestrial climate estimates for the Eocene to Oligocene interval, based on bioclimatic analysis of terrestrially derived spore and pollen assemblages preserved in marine sediments from the Norwegian–Greenland Sea. Our data indicate a cooling of ~5 °C in cold-month (winter) mean temperatures to 0–2 °C, and a concomitant increased seasonality before the Oi-1 glaciation event. These data indicate that a cooling component is indeed incorporated in the $\delta^{18}O$ isotope shift across the Eocene–Oligocene transition. However, the relatively warm summer temperatures at that time mean that continental ice on East Greenland was probably restricted to alpine outlet glaciers.

A rapid two-step shift in benthic foraminiferal $\delta^{18}O$ (Oi-1 glaciation)[1,2] marks a period of dramatic global change, with the development of an Antarctic ice sheet by the Eocene–Oligocene transition, ~33.5 million years (Myr) ago. The proposed driving mechanism for this climate change is the lowering of atmospheric CO_2 from high levels during the Eocene epoch (>1,000 parts per million by volume, p.p.m.v.) to lower levels (~560 p.p.m.v.) in the Oligocene epoch[4–6]. However, it has been argued that the changes in $\delta^{18}O$ composition across this boundary are too large to be explained by ice build-up on Antarctica alone[2], which, combined with continued debate over whether there was ocean cooling associated with the $\delta^{18}O$ shift[3,15–17], raises the possibility of contemporaneous ice development in the Northern Hemisphere. Although recent sedimentological evidence supports the development of some continental ice in the Northern Hemisphere at that time[7,8,18], proxy-based CO_2 estimates suggest that atmospheric CO_2 concentrations were still at least twice the pre-industrial level during the Eocene–Oligocene transition[4,5], which is above the modelled critical threshold for Northern Hemisphere ice development[6,19].

Despite these estimates, there is ample evidence to suggest climate cooling in the late Eocene to Oligocene[11,20,21]. However, there are substantial differences in the extent and timing of the presumed cooling, some records demonstrating little evidence for temperature change[9,10], while other low-resolution records indicate increased seasonality in the Oligocene, shown by a fall in cold-month mean temperatures (CMMTs)[12] and increased aridity[9,10,15,21]. These continental records are mostly from high altitude or mid- to high latitudes, but whereas there is some information on high southern latitude Eocene–Oligocene conditions[22], no comparable information is available for high northern latitudes to enable accurate global climate reconstruction. Moreover, in these high-latitude sequences there is little or no calcareous content, owing to carbonate dissolution, negating the use of foraminiferal $\delta^{18}O$ and Mg/Ca temperature proxies.

Therefore, we examined the acid-insoluble organic fraction of the sediments from three key Ocean Drilling Program sites (ODP sites 643, 985 and 913; Supplementary Fig. S1), producing the first high-resolution terrestrial climate reconstructions using plant palynomorphs (spore and pollen assemblages). In addition, because of limited climate proxy records previously available at these latitudes, we compared our terrestrial climate record with modelling simulations of the mid-Eocene (partial pressure of CO_2, $p_{CO_2} = 1,120$ p.p.m.v.) and the early Oligocene ($p_{CO_2} = 560$ p.p.m.v.).

We constructed a composite section spanning the Eocene to Oligocene by calibrating dinocyst events to the geomagnetic polarity timescale[23], enabling the first palaeoclimatic reconstruction of Greenland through the Eocene and early Oligocene. ODP site 913 (75° 29.356' N, 6° 56.810' E, present water depth ~3,300 m) is the most complete high northern latitude Eocene to Oligocene succession[23]. Spore and pollen assemblages from ODP site 913 are abundant throughout the Eocene sequence (Fig. 1), but the Oligocene is characterized by reworking and reduced recovery (see Supplementary Information). Therefore, this site is supplemented by ODP site 643 (67° 47.11' N, 01° 02.0' E), with a preserved Eocene–Oligocene boundary interval and by ODP site 985 (66° 56.49' N, 6° 27.01' W) with a complete lower Oligocene section (Fig. 2). Despite their different core locations, new modelling data supports the contention that East Greenland is the likely vegetation source area for all sites during the Eocene and Oligocene, allowing climate reconstruction of the Greenland landmass (see Supplementary Information).

The lower Eocene palynofloral assemblage at Site 913B is comprised of taxa typical of extant lowland freshwater swamps on the United States southeast coastal plain[24], namely Taxodiaceae–Cupressaceae (*Taxodium* or *Metasequoia* type), several types of ferns (for example, *Lygodium* and other Schizeaceae), palms/cycads (Arecaceae or Cycadales), *Carya* (Juglandaceae), *Nyssa* (Nyssaceae), and *Salix* (Salicaceae) grains (Fig. 1). The occurrence of palms/cycads in these sediments indicates mesothermal to megathermal conditions and CMMTs over 5 °C (refs 24, 25).

The palynofloral assemblage recovered from middle to upper Eocene sediments from ODP site 913 are consistent with other records for that time[25,26], with characteristic microthermal to mesothermal Pinaceae pollen (for example, *Abies*, *Picea*, *Pinus* and
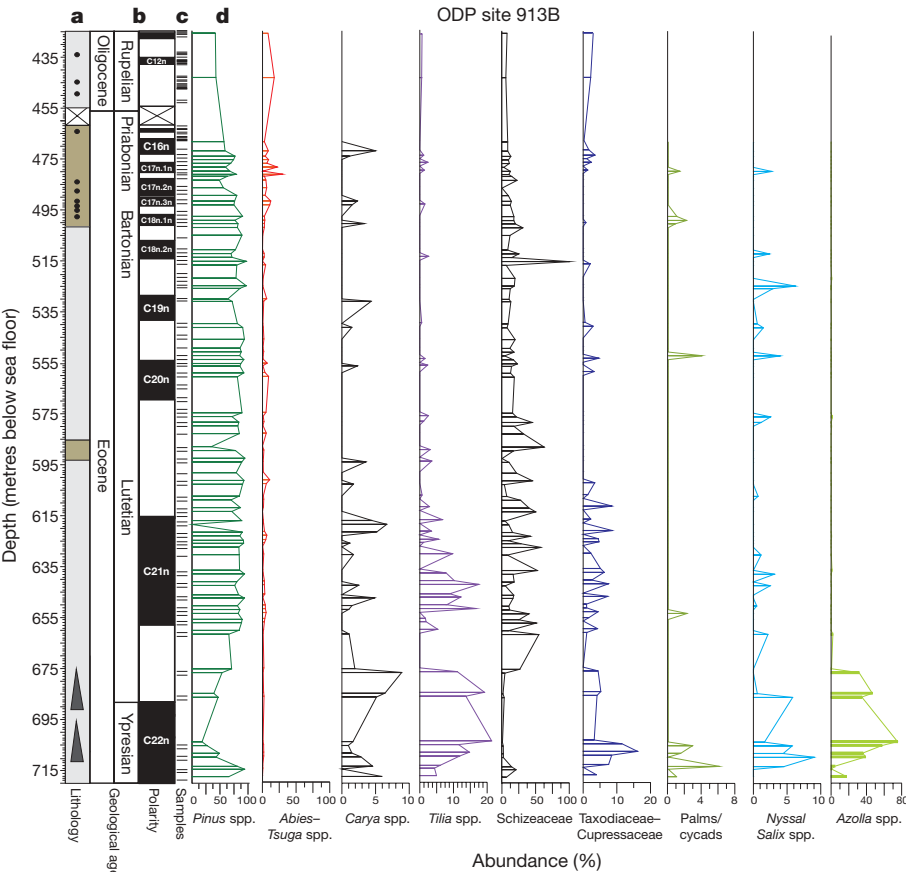
Figure 1 | Data from ODP site 913. a, Lithology of the core, comprising biosiliceous clay (brown), clay and silty clay (light grey), macroscopic dropstones (solid circles), and fining upward sequences (triangles). **b,** Age model. The chronology of ODP site 913 is based on direct correlation with the geomagnetic polarity timescale reported in ref. 23. **c,** Samples analysed for palynology. **d,** Selected pollen and spore percentage abundance.
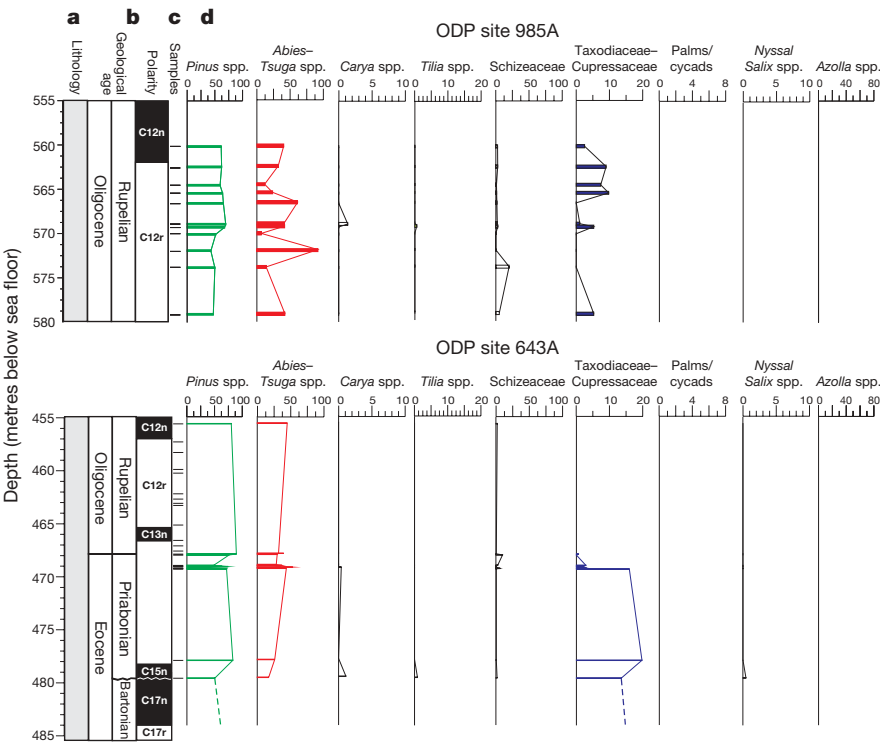

Figure 2 | Data from ODP sites 985 and 643. a, Lithology of the core, comprising clay and silty clay (light grey). **b,** Age models: the chronology for ODP sites 985 and 643 is based on direct correlation to the geomagnetic polarity timescale reported in ref. 23. The chronology for the basal Oligocene for ODP sites 985 and 643 are presented in the Supplementary Information. **c,** Samples analysed for palynology. **d,** Selected pollen and spore percentage abundance.

the rare *Tsuga*) being dominant, indicating extensive coniferous forests on Greenland. Mesothermal conifers such as Taxodiaceae–Cupressaceae and *Sciadopitys*, and dicotyledons such as *Carya*, *Juglans* and *Tilia* are still represented, indicating both allochthonous long-distance pollen transport of preferred wind-dispersed pollen (for example, Pinaceae) from cooler—perhaps mountainous—regions of East Greenland to the marine sediments at Site 913B, combined with shorter-distance transport from the warmer lowland conditions prevailing at the coast. Our climate model indicates that wind speed and direction would not have been sensitive to likely climate changes during the Eocene–Oligocene transition, so we interpret trends in the palynological record (below) as representing changes in the climatic conditions in the source regions and not as changes in atmospheric circulation (Supplementary Information).

Uppermost Eocene–lower Oligocene sediments from the two more southerly ODP sites 643A and 985A (Fig. 1) yield palynofloral assemblages dominated by Pinaceae (*Abies, Picea* and *Pinus*), but with a substantial increase in *Tsuga* and *Abies* relative to the mid-Eocene assemblages of Site 913B (Fig. 2). Moreover, the absence of mesothermal frost-sensitive taxa (that is, palms/cycads, *Engelhardtia* and *Nyssa*) and rarity of other mesothermal elements (such as 'Taxodiaceae', *Tilia, Carya*), suggests the widespread development of an evergreen *Pinus-, Tsuga-* and *Abies*-dominated conifer forest under microthermal conditions by the Oligocene. The shifts in dominance and floral composition documented in the latest Eocene to earliest Oligocene by the pollen sums, although partly confounded by their different core locations (see Supplementary Information), appears to show a displacement of warmer, moderately diverse mixed conifer–broadleaf vegetation and plant types from lowland sites caused by the expansion of a cooler, less-diverse Pinaceae-dominated coniferous forest type.

To derive a quantified palaeoclimatic signal from the pollen and spore assemblages we have applied bioclimatic analysis using the nearest living relatives (NLRs) method[24] (see Supplementary Information) to reconstruct mean annual temperature, mean annual precipitation, CMMT and warm month mean temperature (WMMT) for the Eocene to Oligocene transition (Fig. 3). Our data indicate that the Eocene Greenland climate was generally warm, with mean annual temperature $14\,^{\circ}\mathrm{C} \pm 3\,^{\circ}\mathrm{C}$, CMMT $> 5\,^{\circ}\mathrm{C}$ and WMMT ≈ 18–$24\,^{\circ}\mathrm{C}$, respectively (Fig. 3b–d). A small number of samples in the late Eocene from Site 913B give estimates of CMMT of 4–$5\,^{\circ}\mathrm{C}$ and WMMT $< 20\,^{\circ}\mathrm{C}$. We calculated the first precipitation estimates for the Eocene of Greenland, which are high (mean annual precipitation $>120\,\mathrm{cm\,yr}^{-1}$), although there are large uncertainties in this estimate (Fig. 3e). The mean annual temperature estimates generated are consistent with previous estimates of mid-Eocene climate records from Axel Heiberg Island in the Canadian Arctic further to the west[25]. However, our mid-Eocene CMMT estimates are slightly warmer ($>5\,^{\circ}\mathrm{C}$) than that estimated using leaf physiognomy from macrofloras from Axel Heiberg Island ($\sim 0\,^{\circ}\mathrm{C}$, ref. 25). Cooler winter conditions in the Eocene Canadian Arctic than in Greenland are supported by our model simulations (Fig. 4), but may also reflect the different methodologies used, because leaf physiognomy has been shown to yield cooler estimates than NLR-based approaches for the same fossil assemblage[24,25].

The latest Eocene to early Oligocene climate estimates from ODP sites 643 and 985 have been corrected for palaeolatitude ($0.4\,^{\circ}\mathrm{C}$ per $1\,^{\circ}$ latitude, as presented in ref. 25; see Supplementary Information) so that they can be directly compared with ODP site 913. Our data indicates that for the latest Eocene to Oligocene, mean annual temperature is estimated at 10–$11\,^{\circ}\mathrm{C} \pm 3\,^{\circ}\mathrm{C}$, representing a decline of $\sim 3\,^{\circ}\mathrm{C} \pm 3\,^{\circ}\mathrm{C}$ from the mid- to late Eocene interval. However,
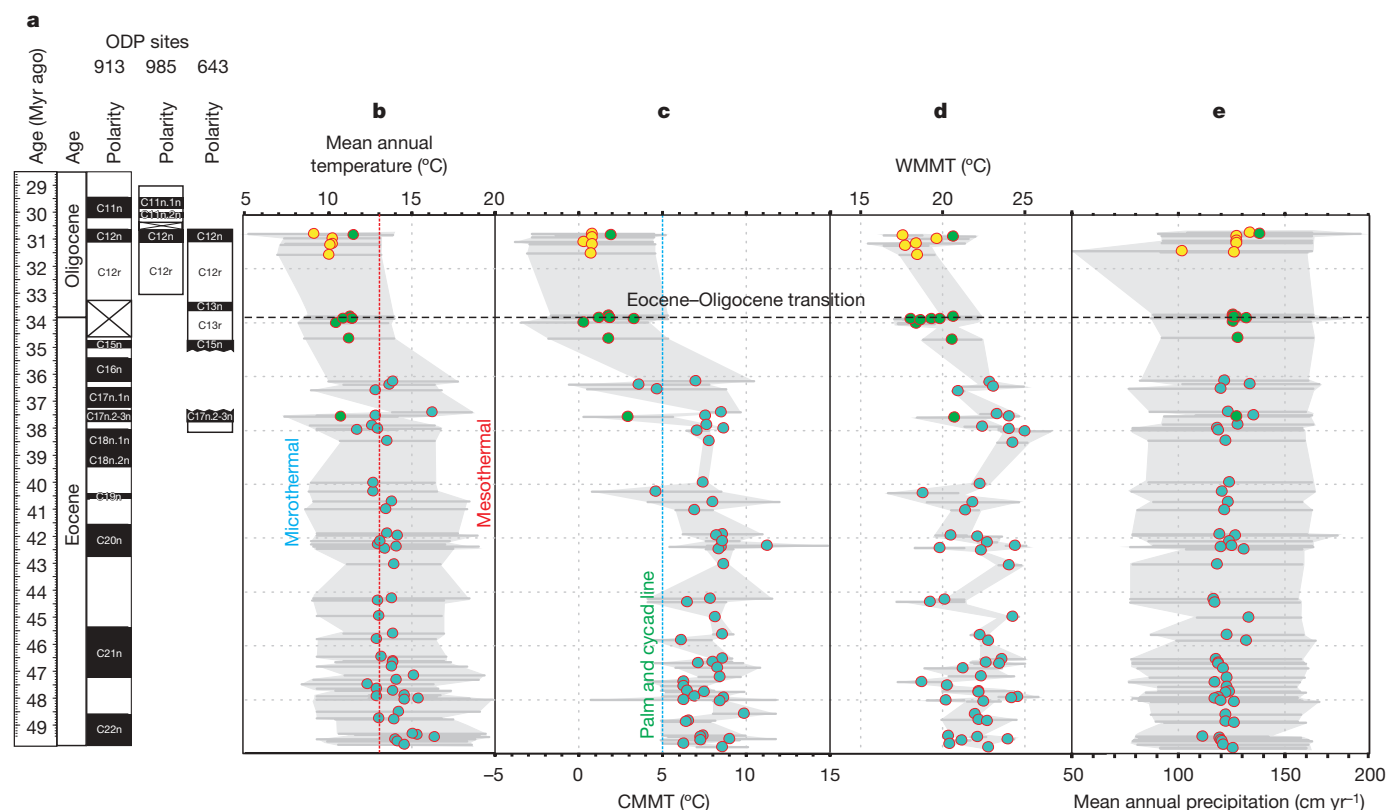

Figure 3 | Data from ODP sites 913, 643 and 985, Norwegian–Greenland Sea. a, Composite chronologies based on direct correlation to the geomagnetic polarity timescale (ref. 23; see Supplementary Information). **b–e**, Composite record of Eocene to Oligocene climate data. **b**, Mean annual temperature. Red dashed line defines transition from mesothermal to microthermal conditions. **c**, CMMT. Blue dashed line represents the temperature at which palms/cycads do not occur, as discussed in ref. 25. **d**, WMMT. **e**, Mean annual precipitation. For all plots, blue circles indicate ODP site 913, yellow circles indicate site 985 and green circles indicate site 643. **b–e**, Horizontal error bars and shaded area represent the minimum and maximum estimate returned from the method (ref. 24).
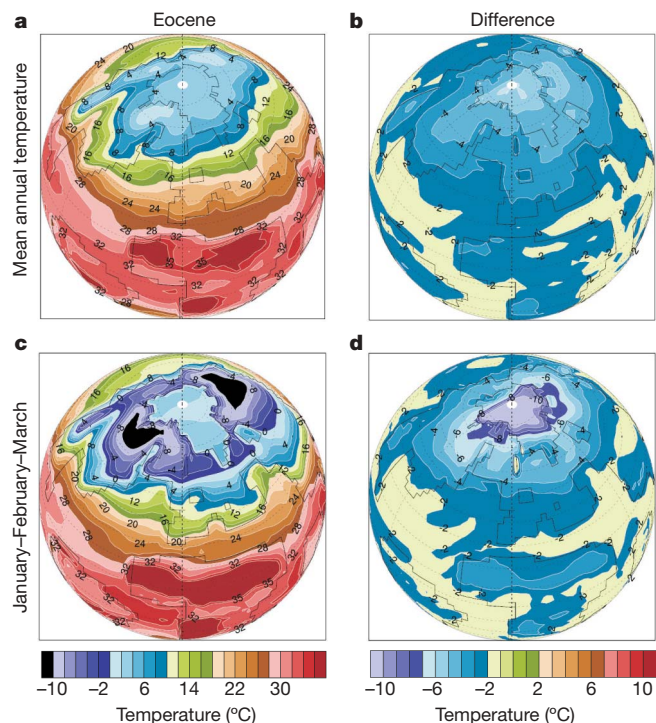
971

Figure 4 | Model surface temperature results (shown in °C) supporting bioclimatic estimates. Left panels show Eocene (p_{CO_2} = 1,120 p.p.m.v.) temperatures. Right panels show temperature differentials between the Eocene and the Oligocene (p_{CO_2} = 1,120–560 p.p.m.v.). **a, b,** Mean annual temperature; **c, d,** January–February–March (representative of CMMT).

although this amount of cooling is consistent with earlier records[27] and our global circulation model simulations (Fig. 4a, b), it is also within the error limits of the data. Our data also indicates a slight decline in WMMT, although this may not be significant because our estimates fall within the error limits and range seen in the Eocene. Nor do we see a significant change in mean annual precipitation (>120 cm yr^{-1}), suggesting continued high precipitation and relatively warm summer temperatures during the earliest Oligocene. There is some indication of cyclical variability in the WMMT mid-Eocene record. However, of greatest significance, winter temperatures (as CMMT) show greater variability than in the mid-Eocene, leading up to the Eocene–Oligocene boundary interval with some estimates of CMMT ~5 °C, and consistently cooler values in the early Oligocene (CMMT ~0–2 °C) than recorded for most of the mid- to late Eocene record (CMMT ~8 °C). The drop in CMMT in the latest Eocene to Oligocene is apparent even without the latitudinal correction.

Our interpretation of cooler winters in the latest Eocene to early Oligocene on Greenland is consistent with other climate records[10,12], while our global circulation model simulations (Fig. 4c, d), indicate that our climate record for Greenland is representative of the northern high latitudes as a whole. Therefore our data indicates that in high northern latitudes the Eocene–Oligocene transition saw a shift towards more extreme seasonal temperature ranges, greater in magnitude than the mean annual temperature changes. The shift to cooler winters is reflected in particular by (1) a loss or dramatic reductions of both freezing-intolerant palms/cycads (plants that today are not found in climates where CMMT < 5 °C, ref. 25) and warm-temperate broadleaf taxa such as hickory (*Carya*), *Engelhardtia*, tupelo (*Nyssa*) and basswood (*Tilia*) by the late Eocene, and (2) by a rise to dominance in the latest Eocene and Oligocene of typical mid-latitude to boreal forest conifers such as fir (*Abies*), hemlock (*Tsuga*) and other Pinaceae. The change in vegetation we report occurs before the Oi-1 glaciation event, and is contemporaneous with the first ice build-up on Greenland ~38 Myr ago[7]. Our climate model results are consistent with this pattern, clearly showing the initiation of thick, extensive

winter sea-ice and thin perennial Arctic sea-ice across the transition interval and a concomitant decrease of winter temperatures of greater magnitude than that in mean annual temperature. Other grain size evidence from ODP site 913 for an earlier inception of ice on Greenland (44 Myr ago, ref. 28) is discounted here, because these deeper sediments contain no macroclasts/dropstones and are characterized by abundant concretions, authigenic quartz grains and possible contourite deposits, all of which point towards a non-glacial origin for these deeper grain-size records.

The climatic estimates presented here provide the first continental temperature constraints for high northern latitudes for the Eocene to Oligocene, and therefore have implications for the ongoing debate concerning Cenozoic ice budgets, the timing of the inception of Northern Hemisphere glaciation[1–3,7,8,16–18,29,30], and our understanding of Cenozoic atmospheric carbon dioxide levels[4–6,19]. The cooler CMMTs and increased seasonality we document occurs before the Cenozoic uplift in Greenland commencing ~36 Myr ago[31], but the contemporary northward movement of Greenland to impinge on the Eurasian basin and the Lomonosov ridge[32] may have influenced Northern Hemisphere ice build-up in the earliest Oligocene, especially at high altitudes. However, the relatively warm summer temperatures at that time means it is likely that Northern Hemisphere ice on East Greenland was restricted to alpine outlet glaciers[3,7,30]. This interpretation is consistent with high-resolution Eocene isotope records[30] and both proxy and model reconstructions of Cenozoic CO_2 (refs 4, 5), which remain well above the critical threshold (~2.8 times the pre-industrial level) for bipolar ice-sheet development[6,19]. The winter temperature decrease is consistent with the initiation of thick, extensive Arctic Ocean winter sea-ice occurring at p_{CO_2} ≈ 560 p.p.m.v. The winter cooling revealed by the data may have increased deep convection and been responsible for the apparent initiation of Northern Component Water formation during this interval[33].

METHODS SUMMARY

Palynology. The palynological techniques used to process the 141 samples are described in ref. 23. Residues were analysed using a stereo-binocular microscope until the entire slide was counted for terrestrial palynomorphs.

Bioclimatic analyses. The method of bioclimatic analysis is essentially the same as described in ref. 24, but with additional data sources enabling more robust and diverse climatic profiles to be developed for each taxon, as described in the online-only Methods. To account for the low species richness found in many samples, we tested the effect of rarefaction of individual species-rich samples on the estimate by recalculating the estimate using successively fewer of the available NLRs with climate profiles. Rarefied samples with fewer than 12 NLRs with climate profiles produced widely divergent estimates at the same species richness, whereas the estimates for rarefied samples with 12 or more NLRs were consistently well within the error of the estimate (see Supplementary Information). Accordingly, only samples with 12 or more NLRs with climate profiles were used in our analysis. Those taxa that were used in this study and their NLRs can be found in Supplementary Table S1.

Climate model simulations. To estimate wind and current directions and provide a physical framework for understanding the climate signal recorded by proxies, we conducted a series of palaeoclimate general circulation model experiments. We used the atmospheric component (CAM version 3) of the National Center for Atmospheric Research, Community Climate System Model (version 3), at a spectral resolution of T42. The sea surface temperatures used to drive these simulations were derived from long, equilibrated, fully coupled, synchronous, unaccelerated simulations of the Eocene (p_{CO_2} = 1,120 p.p.m.v.) and Oligocene (p_{CO_2} = 560 p.p.m.v.). Topography, bathymetry and vegetation distributions and other boundary conditions are appropriate for this time interval and they and the model are described in depth in the online-only Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Zachos, J. C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
2. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).

3.   Lear, C. *et al.* Cooling and ice growth across the Eocene-Oligocene transition. *Geology* **36** (3) 251–354, doi:210.1130/G1124 (2008).

4.   Pagani, M. *et al.* Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).

5.   Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide over the past 60 million years. *Nature* **406**, 695–699 (2000).

6.   DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2. *Nature* **421**, 245–249 (2003).

7.   Eldrett, J., Harding, I. C., Wilson, P., Butler, E. & Roberts, A. Continental ice in Greenland during the Eocene and Oligocene. *Nature* **446**, 176–179 (2007).

8.   Moran, K. *et al.* The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* **441**, 601–605 (2006).

9.   Kohn, M. J. *et al.* Climate stability across the Eocene-Oligocene transition, southern Argentina. *Geology* **32**, 621–624 (2004).

10. Grimes, S. T., Hooker, J. J., Collinson, M. E. & Mattey, D. P. Summer temperatures of late Eocene to early Oligocene freshwaters. *Geology* **33**, 189–192 (2005).

11. Zanazzi, A., Kohn, M. J., MacFadden, B. J. & Terry, D. O. Large temperature drop across the Eocene–Oligocene transition in central North America. *Nature* **445**, 639–642 (2007).

12. Ivany, L. C., Patterson, W. P. & Lohmann, K. C. Cooler winters as a possible cause of mass extinctions at the Eocene/Oligocene boundary. *Nature* **407**, 887–890 (2000).

13. Terry, D. O. Jr. Paleopedology of the Chadron Formation of northwestern Nebraska; implications for paleoclimatic change in the North American Midcontinent across the Eocene-Oligocene boundary. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **168**, 1–38 (2001).

14. Sheldon, N. D., Retallack, G. J. & Tanaka, S. Geochemical climofunctions from North American soils and application to paleosols across the Eocene-Oligocene boundary in Oregon. *J. Geol.* **110**, 687–696 (2002).

15. Dupont-Nivet, G. *et al.* Tibetan Plateau aridification linked to global cooling at the Eocene–Oligocene transition. *Nature* **445**, 635–638 (2007).

16. Lear, C. H., Rosenthal, Y., Coxall, H. K. & Wilson, P. A. Late Eocene to early Miocene ice-sheet dynamics and the global carbon cycle. *Paleoceanography* **19**, PA4015, doi:10.1029/2004PA001039 (2004).

17. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).

18. St John, K. Cenozoic ice-rafting history of the Central Arctic Ocean: terrigenous sands on the Lomonosov ridge. *Paleoceanography* **23**, PA1S05 (2008).

19. DeConto, R. *et al.* Thresholds for Cenozoic bipolar glaciation. *Nature* **455**, 652–656 (2008).

20. Kobashi, T., Grossman, E. L., Yancey, T. E. & Dockery, D. T. III. Reevaluation of conflicting Eocene tropical temperature estimates; molluskan oxygen isotope evidence for warm low latitudes. *Geology* **29**, 983–986 (2001).

21. Retallack, G. J. *et al.* Eocene-Oligocene extinction and paleoclimatic change near Eugene, Oregon. *Geol. Soc. Am. Bull.* **116**, 817–839 (2004).

22. Thorn, V. C. & DeConto, R. M. Antarctic climate at the Eocene/Oligocene boundary; climate model sensitivity to high latitude vegetation type and comparisons with the palaeobotanical record. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 134–157 (2006).

23. Eldrett, J. S., Harding, I. C., Firth, J. V. & Roberts, A. P. Magnetostratigraphic calibration of Eocene-Oligocene dinoflagellate cyst biostratigraphy from the Norwegian-Greenland Sea. *Mar. Geol.* **204**, 91–127 (2004).

24. Greenwood, D. R., Archibald, S. B., Mathewes, R. W. & Moss, P. T. Fossil biotas from the Okanagan Highlands, southern British Columbia and northeastern Washington State: climates and ecosystems across an Eocene landscape. *Can. J. Earth Sci.* **42**, 167–185 (2005).

25. Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature-gradients. *Geology* **23**, 1044–1048 (1995).

26. Norris, G. Spore-pollen evidence for early Oligocene high-latitude cool climatic episode in northern Canada. *Nature* **297**, 387–389 (1982).

27. Schouten, S. *et al.* Onset of long-term cooling of Greenland near the Eocene-Oligocene boundary as revealed by branched tetraether lipids. *Geology* **36**, 147–150, doi:10.1130/G24332A.1 (2008).

28. Tripati, A. *et al.* Evidence for Northern Hemisphere glaciation back to 44 Ma from ice-rafted debris in the Greenland Sea. *Earth Planet. Sci. Lett.* **265**, 112–122 (2008).

29. Billups, K. & Schrag, D. P. Application of benthic foraminiferal Mg/Ca ratios to questions of Cenozoic climate change. *Earth Planet. Sci. Lett.* **209**, 181–195, doi:10.1016/S0012–821X(03)00067–0 (2003).

30. Edgar, K. M., Wilson, P. A., Sexton, P. F. & Suganuma, Y. No extreme bipolar glaciation during the main Eocene calcite compensation shift. *Nature* **448**, 908–911 (2007).

31. Bonow, J. M., Japsen, P., Lidmar-Bergstrom, K., Chalmers, J. A. & Pedersen, A. K. Cenozoic uplift of Nuussuaq and Disko, West Greenland—elevated erosion surfaces as uplift markers of a passive margin. *Geomorphology* **80**, 325–337 (2006).

32. Brozena, J. M. *et al.* New aerogeophysical study of the Eurasia Basin and Lomonosov Ridge: implications for basin development. *Geology* **31**, 825–828 (2003).

33. Via, R. K. & Thomas, D. J. Evolution of Atlantic thermohaline circulation: Early Oligocene onset of deep-water production in the North Atlantic. *Geology* **34**, 441–444 (2006).

Author Contributions J.S.E. and I.C.H. conceived the project, collected samples and data, and completed the palynological interpretation. D.R.G. and J.S.E. completed the bioclimatic analysis of the spore–pollen record, and led the write-up. M.H. conducted all numerical climate modelling and Lagrangian trajectory analysis. All authors were involved in developing the palaeoclimate interpretation in this work

nature

METHODS

Bioclimatic analyses. The method of bioclimatic analysis is essentially the same as the coexistence approach of ref. 34, and requires the development of 'climatic profiles' for living taxa, based on the climatic envelopes controlling the distributions of modern plant genera. A library of climatic profiles is then produced for individual taxa, based on climatic values such as mean annual temperature, mean annual precipitation, and CMMT[35,36]. The climatic profiles used in this study were from four sources: (1) a study of North American tree genera[36], (2) the Palaeoflora database[37] (and personal communication T. Utescher, Bonn Univ.), which includes a wider range of extant plant genera than the North American database and also includes European and Asian distribution records (for example, *Engelhardtia*, *Pterocarya*), (3) Natural Resources Canada[38], used for North American herbaceous and shrub taxa not present in the other databases (for example, *Lonicera*) and (4) the Australian National Herbarium database[39], which was analysed using the ANUCLIM software[35], and used for tropical and temperate taxa not represented in the other databases (for example, *Gleichenia*). Where values were available for the same taxon from more than one source, sources (1) and (2) were used preferentially, or the values with the widest range were used (for example, coolest minimum and warmest maximum for CMMT).

Climate model simulations. The NCAR Community Climate System Model version 3 (CCSM3)[40] is a coupled climate model with separate components that can be integrated separately or in tandem for the atmosphere (CAM), land surface/vegetation (CLM), sea ice (CSIM), and ocean (POP). For the purposes of this study relatively high atmospheric resolution was considered beneficial (because the relevant Arctic land-sea distributions are not well resolved at T31, a typical palaeoclimate general circulation model resolution) so we performed stand-alone atmospheric general circulation model simulations using CAM3 (ref. 41) at T42 resolution driven by sea surface temperatures derived from Eocene–Oligocene fully coupled CCSM3 simulations. The coupled simulations were carried out at the standard palaeoclimate resolution parameters for CCSM3, with T31 resolution in the atmospheric model (CAM3), and a nominal resolution of $\times 3'$ in the ocean[42,43]. These fully coupled simulations were carried out for over 1,500 years and are fully equilibrated at a range of CO_2 values as described in ref. 44. The global mean sensitivity of the model is about 2.2 °C per doubling, both for these simulations and for modern to doubled CO_2 concentrations. This is the normal range of sensitivity for the CCSM3 model[45]. The two CAM3 stand-alone atmospheric general circulation model simulations used in this study are performed with p_{CO_2} values of 1,120 and 560 p.p.m.v. and driven by sea surface temperatures and sea-ice fractions from the coupled simulations with the same CO_2. All other parameters were held constant. In all cases, vegetation, orography and topography, and other parameters were based on reconstructed Eocene distributions[46–48].

Accordingly, to estimate palaeo-wind transports, and surface temperatures we used the T42 CAM3 simulations. The simulations are fully equilibrated: the top-of-atmosphere radiative imbalance is -0.047 W m^{-2} and -0.004 W m^{-2} for the simulated low and high p_{CO_2} concentrations respectively; global mean terrestrial temperatures reached quasi-steady state after 5 years and the simulations were continued for at least an additional 60 years. Means from the last 15 years are

discussed. All differences discussed are statistically significant at the 95% level based on a Student's *t*-test.

To determine the vegetation source region and the influence of atmospheric transportation of spore–pollen transport pathways we performed a fully Lagrangian tracer investigation for this region. Lagrangian back-trajectories (refs 49, 50) were calculated centred on the Greenland Sea and followed back for ten days using the six-hourly low-level wind fields. The trajectories were started every ten days and for each year, and this was repeated for ten years to produce a meaningful ensemble of back-trajectories (that is, approximately 360 separate releases of a puff of passive tracer particles). The ensemble mean distribution should be considered as an approximate probability distribution function for the likelihood that any given region is a potential source is shown in Supplementary Figs S4–S6.

34. Utescher, T., Mosbrugger, V. & Ashraf, A. R. Terrestrial climate evolution in northwest Germany over the last 25 million years. *Palaios* **15**, 430–449 (2000).
35. Greenwood, D. R., Archibald, S. B., Mathewes, R. W. & Moss, P. T. Fossil biotas from the Okanagan Highlands, southern British Columbia and northeastern Washington State: climates and ecosystems across an Eocene landscape. *Can. J. Earth Sci.* **42**, 167–185 (2005).
36. Thompson, R. S., Anderson, K. H. & Bartlein, P. J. Atlas of relations between climatic parameters and distributions of important trees and shrubs in North America. *US Geol. Surv. Prof. Pap.* 1650, A & B (USGS, 1999).
37. Palaeoflora Database at ⟨http://www.geologie.uni-bonn.de/Palaeoflora/Palaeoflora_home.htm⟩.
38. Natural Resources Canada at ⟨http://planthardiness.gc.ca/ph_main.pl?lang=en⟩.
39. Integrated. Botanical Information System (IBIS) Australian National Herbarium at ⟨http://www.anbg.gov.au/cgi-bin/anhsir⟩.
40. Collins, W. D. *et al.* The Community Climate System Model Version 3 (CCSM3). *J. Clim.* **19**, 2122–2143 (2006).
41. Collins, W. D. *et al.* The formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3). *J. Clim.* **19**, 2144–2161 (2006b).
42. Yeager, S., Shields, C., Large, W. & Hack, J. J. The low resolution CCSM3. *J. Clim.* **19**, 2545–2566 (2006).
43. Kiehl, J. T. & Shields, C. A. Climate simulation of the latest Permian: implications for mass extinction. *Geology* **33**, 757–760 (2005).
44. Liu, Z. *et al.* Global cooling during the Eocene-Oligocene climate transition. *Science* **323**, 1187–1190 (2009).
45. Kiehl, J. T., Shields, C. A., Hack, J. J. & Collins, W. D. The climate sensitivity of the Community Climate System Model Version 3 (CCSM3). *J. Clim.* **19**, 2584–2596 (2006).
46. Sewall, J. O., Sloan, L. C., Huber, M. & Wing, S. Climate sensitivity to changes in land surface characteristics. *Glob. Planet. Change* **26**, 445–465 (2000).
47. Huber, M. & Sloan, L. C. Heat transport, deep waters, and thermal gradients: Coupled simulation of an Eocene ''greenhouse'' climate. *Geophys. Res. Lett.* **28**, 3481–3484 (2001).
48. Huber, M., Sloan, L. C. & Shellito, C. in *Causes and Consequences of Globally Warm Climates in the Early Paleogene* (eds Wing, S. L., Gingerich, P. D., Schmitz, B. & Thomas, E.) *Geol. Soc. Am. Spec. Pap.* **369**, 25–47 (2003).
49. Yang, H. & Pierrehumbert, R. T. Production of dry air by isentropic mixing. *J. Atmos. Sci.* **51**, 3437–3454 (1994).
50. Huber, M., McWilliams, J. C. & Ghil, M. A climatology of turbulent dispersion in the troposphere. *J. Atmos. Sci.* **58**, 2377–2394 (2001).

Malestrom

LETTERS

Creep cavitation can establish a dynamic granular fluid pump in ductile shear zones

F. Fusseis[1], K. Regenauer-Lieb[1,2], J. Liu[2], R. M. Hough[2] & F. De Carlo[3]

The feedback between fluid migration and rock deformation in mid-crustal shear zones is acknowledged as being critical for earthquake nucleation, the initiation of subduction zones and the formation of mineral deposits[1–3]. The importance of this poorly understood feedback is further highlighted by evidence for shear-zone-controlled advective flow of fluids in the ductile lower crust[4] and the recognition that deformation-induced grain-scale porosity is a key to large-scale geodynamics[5,6]. Fluid migration in the middle crust cannot be explained in terms of classical concepts. The environment is considered too hot for a dynamic fracture-sustained permeability as in the upper crust[7], and fluid pathways are generally too deformed to be controlled by equilibrium wetting angles that apply to hotter, deeper environments[8–10]. Here we present evidence that mechanical and chemical potentials control a syndeformational porosity generation in mid-crustal shear zones. High-resolution synchrotron X-ray tomography and scanning electron microscopy observations allow us to formulate a model for fluid migration in shear zones where a permeable porosity is dynamically created by viscous grain-boundary sliding, creep cavitation, dissolution and precipitation. We propose that syndeformational fluid migration in our 'granular fluid pump' model is a self-sustained process controlled by the explicit role of the rate of entropy production of the underlying irreversible mechanical and chemical microprocesses. The model explains fluid transfer through the middle crust, where strain localization in the creep regime is required for plate tectonics, the formation of giant ore deposits, mantle degassing and earthquake nucleation.

Our findings provide a key component for the understanding of creep instabilities in the middle crust.

The feedback between fluid migration and ductile deformation can be studied by investigating strain gradients and the associated changes in porosity in exhumed shear zones[11]. We sampled the margin and centre of a metre-scale shear zone that forms part of the Redbank shear zone in central Australia, a major tectonic element in the region[12]. The particular segment (UTM coordinates 53K 357142E, 7390093S) deformed at upper-greenschist-facies metamorphic conditions (350–550 °C; lithostatic pressure, ~500 MPa)[12]. The sample covers a strain gradient expressed in the reorientation and overprint of a pre-existing foliation as well as a significant grain-size reduction towards the shear-zone centre (Fig. 1a). The grain-size reduction from macroscopically resolvable grains to a 'glassy' mylonite indicates strain-dependent softening[13]. We therefore assume that the fine-grained mylonite in the shear zone centre was mechanically softer than the coarser-grained marginal segments (Fig. 1a). This implies that, as the mylonite formed during shear-zone growth, the margins were progressively deactivated. This allows the strain gradient to be interpreted as a temporal sequence[11] and, therefore, the investigation of progressive mylonite formation and any associated change in the porosity distribution and fluid migration.

We used synchrotron X-ray microtomography[14] to visualize all pores larger than 1.3 µm in diameter in 55 subsamples covering the entire shear-zone margin (Supplementary Fig. 1). Here we present clear evidence that the porosity architecture evolved with the mylonite. The total porosity in marginal, low-strain domains of the hand
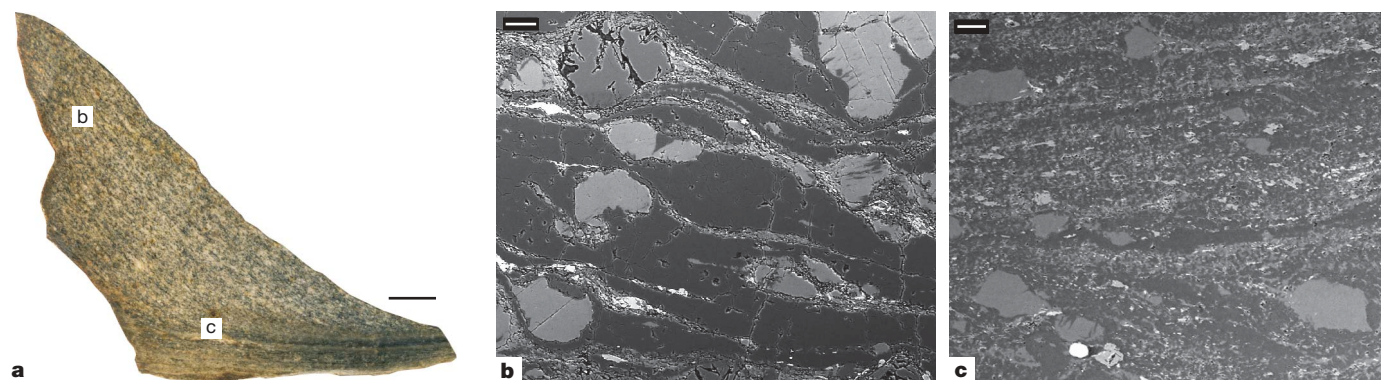

Figure 1 | The investigated hand specimen from a mid-crustal shear zone, and the microstructural evolution with increasing shear strain. a, The hand specimen exhibits a marked strain gradient towards the shear-zone centre (from top to bottom). Scale bar, 2 cm. Locations 'b' and 'c' are shown in **b** and **c**, respectively. **b**, Low-strain fabric from the shear-zone margin showing quartz (dark grey), plagioclase clasts (intermediate grey) and K-feldspar clasts (light grey) as well as epidote (white). We note the fine-grained multiphase domains extending from the feldspar clasts. Scale bar, 100 µm. **c**, High-strain microfabric from the shear-zone centre showing quartz (dark grey), K-feldspar as clasts and in fine-grained matrix (light grey) and epidote (lightest grey). Scale bar, 100 µm.

[1]School of Earth & Environment, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia. [2]CSIRO Exploration & Mining, 26 Dick Perry Avenue, Kensington, Western Australia 6151, Australia. [3]Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.
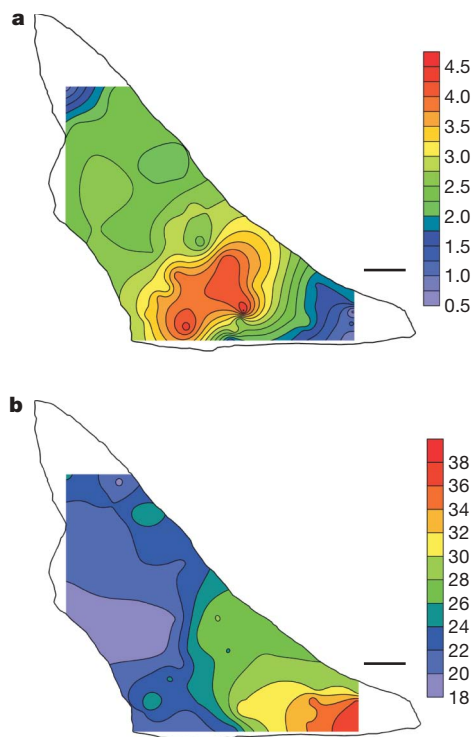
a

b

Figure 2 | Porosity distribution across a strain gradient in the investigated mid-crustal shear zone. a, Contoured percentage total porosity (colour scale) in hand specimen shown in Fig. 1a. **b,** Percentage relative frequency of smallest pores (1.3–3.6 μm; colour scale) in hand specimen shown in Fig. 1a. Scale bars, 2 cm.

specimen is on the order of 2.5% and increases to more than 4.5% towards the shear-zone centre (Fig. 2a). Over the same distance, the relative frequency of small pores between 1.3 and 3.9 μm in diameter increases from below 25% to more than 35% (Fig. 2b). Smaller pores (<1.3 μm in diameter) that were imaged with a field-emission scanning electron microscope (FESEM) are not captured in Fig. 2.

Our FESEM work on rock thin sections and split samples linked the porosity evolution to the microstructural evolution of the rock. The latter is characterized by the transformation of the pre-existing gneissic microfabric into an ultramylonitic one: in marginal low-strain domains, fabrics are dominated by K-feldspar and plagioclase clasts interlayered with quartz ribbon bands and thin seams of fine-grained mixtures of secondary K-feldspar, muscovite, quartz, biotite and epidote (Fig. 1b). Towards the shear-zone centre, this multiphase mineral mixture increases in proportion and gradually forms an interconnected matrix whereas quartz in the ribbon bands recrystallizes dynamically by subgrain rotation and grain-boundary bulging. In the shear-zone centre, the microfabric comprises a few small K-feldspar clasts floating in the multiphase matrix, with minor quartz-enriched domains representing former ribbon bands (Fig. 1c). Plagioclase disappears with increasing strain. We interpret the feldspar clasts to have decayed by hydration reactions[15,16], with the reaction products forming the multiphase matrix. However, the clasts are rarely directly surrounded by the secondary minerals. The occurrence of small secondary reaction products (mostly K-feldspar and mica) in quartz ribbon bands indicates that the chemical components necessary to form them were redistributed by a fluid phase that required a permeable porosity (Supplementary. Fig. 2).

The hydration reaction and material redistribution are reflected by a change of porosity architecture with increasing strain and the evidence for dissolution preserved in all but the smallest pore shapes. As can be seen in three-dimensional visualizations (Supplementary Movie 1), porosity in the less deformed parts of the sample is dominated by comparatively large pores forming clusters within decaying

plagioclase clasts. These intragranular pores are mostly flake-shaped and form skeleton-like features sometimes more than 100 μm long and tens of micrometres wide (Supplementary Movie 1). FESEM on split rock samples indicate that the flakes' regular orientation is crystallographically controlled. The corrugated outlines of the flakes suggest that the plagioclase was dissolved in a fluid during the hydration reaction.

Towards the shear-zone centre, most porosity is comprised by high concentrations of very small (micrometre-sized) pores forming pore sheets (Supplementary Movie 2 and Fig. 3a). FESEM and the synchrotron data show that the pore sheets are concentrated in the fine-grained multiphase layers along both phase and grain boundaries (Fig. 3a). In quartz ribbon bands, grain-boundary pores are often found next to very small secondary K-feldspar grains (Supplementary Fig. 2). Grain-boundary pores are usually slit-like or resemble narrow voids with openings of <1 μm, often extending along two or three grain boundaries and even bifurcating at triple junctions (Fig. 3b–d). The fact that they often show a mismatch across opposing faces and lobes extending into neighbouring grains indicates that their shapes were controlled by dissolution. This interpretation is supported by the pockmarked (with individual marks being on the order of 10–20 nm in size) and incised morphology of pore surfaces (Fig. 3b, c).

Synkinematic material redistribution by a fluid phase is also evidenced by the precipitation of neoformed minerals in grain-boundary pores. Many pore surfaces are covered by crystallites and amorphous films (Fig. 3c and arrow in Fig. 3d); however, owing to their nanometre size, these were difficult to identify. Micrometre-sized secondary phases that were found associated with grain-boundary pores in the quartz ribbon bands (Supplementary Fig. 2) were identified as K-feldspar, muscovite and epidote, minerals that result from the feldspar hydration reactions. Their heterogeneous nucleation in grain-boundary pores in the quartz ribbon bands has been interpreted as evidence questioning the dominance of dislocation creep mechanisms and instead favouring diffusion creep[17].

The conclusions we draw from the data presented above follow ref. 12, in which it was suggested that when the multiphase matrix formed an interconnected framework during the progressive microstructural evolution, a transition in the rate-controlling deformation mechanisms occurred. Whereas deformation at the low-strain shear-zone margins was mostly controlled by dislocation creep in the quartz ribbon bands, deformation in the shear-zone centre was mainly accommodated by diffusion creep in the multiphase matrix. Our data indicate that diffusion creep involved a combination of viscous grain-boundary sliding (VGBS) and dissolution or precipitation. The transition to deformation mechanisms that affect mostly grain boundaries coincides with the observed increase in the relative frequency of grain-boundary pores (see Figs 1 and 2 and Supplementary Movies 1 and 2). Because a significant post-deformational modification of pore space seems unlikely (Supplementary Information, section 1), our synchrotron X-ray microtomography and FESEM data suggest that the grain-boundary porosity in the shear-zone centre was created synkinematically by viscous grain-boundary sliding and dissolution. VGBS is well known to create pores in metals and ceramics[18,19] by a process called creep cavitation. During creep cavitation, pores form along grain boundaries, where grains slide with respect to each other[18], and around small heterogeneities in the stress field[19] (Fig. 4). This mechanism was postulated for minerals and rocks[20,21]; however, high crustal pressures were previously believed to rule out the concept for crustal shear zones[22]. This is in contrast to recent observations from natural shear zones and rock deformation experiments that were explained by creep cavitation[23–25].

The evidence for interplay between VGBS, creep cavitation and dissolution porosity in an environment of obvious material redistribution outlines a previously unknown mechanism for fluid migration in mid-crustal shear zones that is illustrated in Fig. 4. The figure
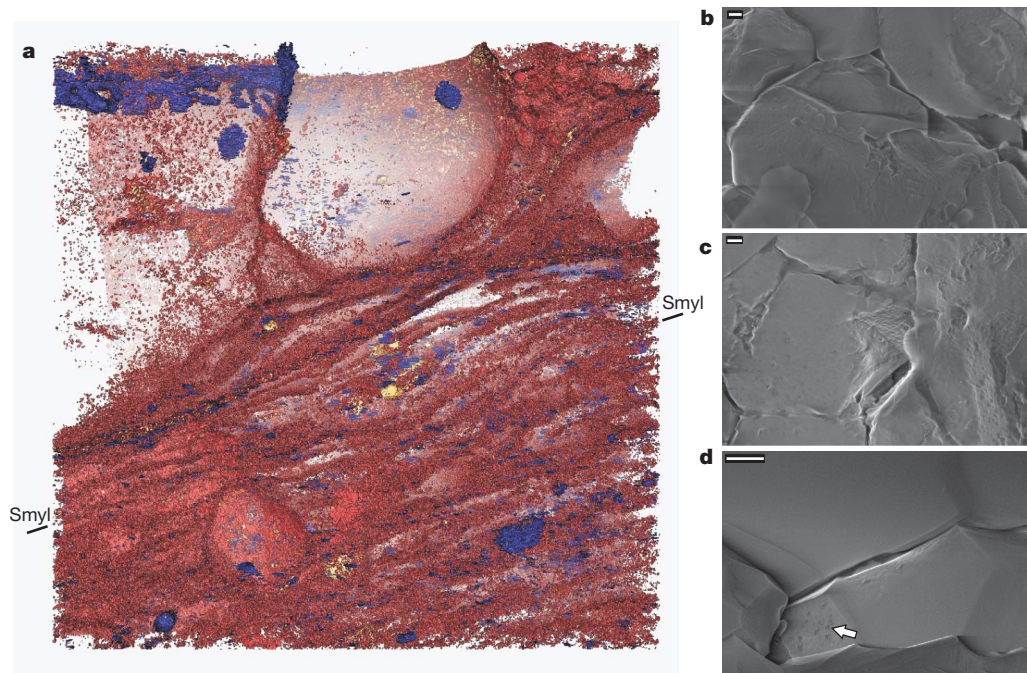
Figure 3 | Three-dimensional visualization of porosity from the shear-zone centre, and grain boundary pores in quartz/K-feldspar mixtures.
a, Micrometre-scale distribution of porosity (red), mica (blue) and oxides (yellow) in subsample H03. We note that pores form pore sheets parallel to the mylonitic foliation (indicated by markers 'Smyl'). The foliation wraps around large clasts of K-feldspar. Between the pore sheets are relatively monomineralic quartz layers, which are pore deficient. The cube shown has a volume of $1.6 \times 1.6 \times 0.8$ mm^3. **b**, Cavity at a triple junction and evidence for material dissolution. Scale bar, 200 nm. **c**, Grain-boundary pore showing a pockmarked surface. We note material redistributed along the grain boundary at the top left of the picture. Scale bar, 200 nm. **d**, Grain-boundary pore. We note microcrystallites precipitated on the uneven grain boundary (arrow at bottom left of picture). Scale bar, 1 μm. **b–d** were photographed in secondary electron mode in split samples.

shows a simplified two-dimensional array roughly representative for the microfabrics in the shear-zone centre. As a function of the differential motion of individual grains during VGBS, creep cavities open up while others close. On the basis of the dynamic grain boundary diffusion experiments in ref. 26, we interpret this to mean that the fluid-filled pores can communicate through grain boundaries during deformation. Thus, grain-boundary sliding and the associated opening and closing of grain-boundary pores give rise to grain-scale fluid pressure differences between neighbouring pores, initiating a granular fluid pump.

This concept of a granular fluid pump is based on the classical assumption of equilibrium between chemical and mechanical energies[8,10,27], with the additional constraint that the rates of change of

Figure 4 | Granular fluid pump model for the VGBS mylonite in the shear-zone centre. The motion vectors (black) illustrate the differential motion of grains. The different pressures in opening and closing pores are colour coded. See text for details. **b, c**, Mechanisms of cavity formation on grain boundaries and triple junctions (locations respectively marked 'b' and 'c' in **a**): cavitation at triple junctions[19] (**b**); cavitation at dislocation tangles near grain boundaries[20] (**c**).

these energies balance during shearing. In thermodynamic terms, this implies that this synkinematic permeability is controlled by the rate of entropy production[28]. The fluid pressure in the individual grain-boundary pores is a composite of the mechanical pumping term and a pressure caused by the solution or precipitation reactions and associated volume changes (Supplementary Information, section 2). The total power dissipated during the process is the rate of change of the pressure work done by the granular fluid pump minus the rate of change of the energy stored in the pore surface. Pore growth is stable if the rate of dissipation is maximized and constant[28]. The granular fluid pump relies on the opening and closing of pores. Hence, our model describes a dynamic permeability that should occur wherever and as long as VGBS operates in rocks with a free fluid phase. It provides steady, non-episodic fluid transfer on the timescales defined by diffusion creep.

The granular fluid pump model solves the inherent problem of high lithostatic pressures by considering pore fluid pressure as a quantity that varies on the grain scale and integrating dissolution and advective fluid transport[7]. The dynamic nature of porosity in the shear-zone centre implies that the pore fluid pressure in opening cavities is controlled by advective fluid flow into the pore, the pore geometry in combination with the surface energy of the fluid, and the pressure created by chemical reactions between fluid and rock. A dynamic equilibrium between lithostatic pressure, pore fluid pressure and pressure from chemical reactions is achieved if, in addition to dissolution and precipitation, an aqueous fluid is advectively transferred into these cavities from nearby collapsing pores (Supplementary Information, section 2). In this way, the pore fluid pressure on the grain scale is close to lithostatic; however, there is no need for episodic large-scale fluid pressures as in previous models[29].

Our results indicate that the interaction of solution or precipitation with creep cavitation during VGBS in crustal shear zones may be expected in many water-saturated mid-crustal shear zones. This has wide implications. Classified as a high-temperature fracture mechanism, creep cavitation leads to localized failure where a critical
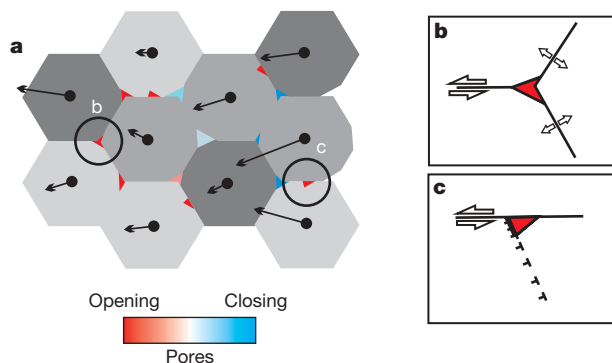
cavity density is exceeded[25,30] (compare with Supplementary Fig. 2d). Applying our findings to phenomena such as accelerated creep events during the earthquake cycle and slow earthquakes may improve the understanding of some aspects of seismic events in the future.

METHODS SUMMARY

High-resolution synchrotron X-radiation microtomography data of 55 subsamples distributed evenly across the hand specimen were collected (Supplementary Fig. 1) at the bending magnet beamline (2-BM) at the Advanced Photon Source, Argonne National Laboratory, Chicago[14]. A Si(111) monochromator provided 20-keV X-rays, images were collected in transmission mode by a charge-coupled-device camera positioned behind the sample in the hutch configuration. The samples are automatically mounted on the rotary stage and placed in the beam by a robot. Data were collected through rotating the samples in $0.25°$ steps over $180°$. The minimum effective pixel size achieved was $1.3\,\mu m$ to give $1.3 \times 1.3 \times 1.3\,\mu m^3$ voxels for three-dimensional rendering of volumes. Different energies were tested to ensure effective transmission through the samples. Isosurfaces of porosity were isolated and visualized from the three-dimensional tomographic data using the software AVIZO.

A Zeiss 1555 FESEM was used to obtain images from thin sections and split samples in secondary and back-scattered electron imaging modes at acceleration voltages between 1 and 30 kV and working distances between 3 and 15 mm. The variable-pressure mode of the microscope allowed us to investigate uncoated samples to resolve details on the nanometre scale.

The porosity values shown in Fig. 2 were determined from 3–5 image slices through each of the 55 subsamples. The pores were thresholded in each of the images and the porosity determined using IMAGE J (http://rsbweb.nih.gov/ij/). The porosity contours in Fig. 2 were constructed from arithmetic means calculated for each of the 55 subsamples.

In Supplementary Information, section 3, we describe the percolation analysis performed on the three-dimensional porosity data.

1. Sibson, R. Crustal stress, faulting and fluid flow. *Spec. Publ. Geol. Soc. (Lond.)* **78**, 69–84 (1994).
2. Regenauer-Lieb, K., Yuen, D. A. & Branlund, J. The initiation of subduction: criticality by addition of water? *Science* **294**, 578–580 (2001).
3. Kolb, J. The role of fluids in partitioning brittle deformation and ductile creep in auriferous shear zones between 500 and 700 °C. *Tectonophysics* **446**, 1–15 (2008).
4. Mack Kennedey, B. & van Soest, M. C. Flow of mantle fluids through the ductile lower crust. *Science* **318**, 1433–1436 (2007).
5. Regenauer-Lieb, K. Dilatant plasticity applied to Alpine collision: ductile void growth in the intraplate area beneath the Eifel volcanic field. *Geodynamics* **27**, 1–21 (1999).
6. Bercovici, D. Generation of plate tectonics from lithosphere-mantle flow and void-volatile self-lubrication. *Earth Planet. Sci. Lett.* **154**, 139–151 (1998).
7. Cox, S. F. & Etheridge, M. A. Coupled grain-scale dilatancy and mass transfer during deformation at high fluid pressures: examples from Mount Lyell, Tasmania. *J. Struct. Geol.* **11**, 147–162 (1989).
8. Watson, E. B. & Brenan, J. M. Fluids in the lithosphere. 1. Experimentally-determined wetting characteristics of CO_2–H_2O fluids and their implications for fluid transport, host-rock physical properties, and fluid inclusion formation. *Earth Planet. Sci. Lett.* **85**, 497–515 (1987).
9. Holness, M. How melted rocks migrate. *Science* **314**, 934–935 (2006).
10. Kohlstedt, D. L. & Holtzman, B. K. Shearing melt out of the Earth: an experimentalist's perspective on the influence of deformation on melt extraction. *Annu. Rev. Earth Planet. Sci.* **37**, 1–33 (2009).
11. Fusseis, F. & Handy, M. R. Micromechanisms of shear zone propagation at the brittle-ductile transition. *J. Struct. Geol.* **30**, 1242–1253 (2008).
12. Fliervoet, T., White, S. H. & Drury, M. R. Evidence for dominant grain-boundary sliding deformation in greenschist- and amphibolite-grade polymineralic ultramylonites from the Redbank Deformed Zone, Central Australia. *J. Struct. Geology* **19**, 1495–1520 (1997).
13. White, S. *et al.* On mylonites in ductile shear zones. *J. Struct. Geol.* **2**, 175–187 (1980).
14. Wang, Y. *et al.* A high-throughput X-ray microtomography system at the Advanced Photon Source. *Rev. Sci. Instrum.* **72**, 2062–2068 (2001).
15. O'Hara, K. & Blackburn, W. H. Volume-loss model for trace-element enrichments in mylonites. *Geology* **17**, 524–527 (1989).
16. Wintsch, R. P. & Yi, K. Dissolution and replacement creep: a significant deformation mechanism in mid-crustal rocks. *J. Struct. Geol.* **24**, 1179–1193 (2002).
17. Kruse, R. & Stünitz, H. Deformation mechanism and phase distribution in mafic high-temperature mylonites from the Jotun Nappe, southern Norway. *Tectonophysics* **303**, 223–249 (1999).
18. Blanchard, C. R. & Chan, K. S. Evidence of grain boundary-sliding-induced cavitation in ceramics under compression. *J. Am. Ceram. Soc.* **76**, 1651–1660 (1993).
19. Kassner, M. E. & Hayes, T. A. Creep cavitation in metals. *Int. J. Plast.* **19**, 1715–1748 (2003).
20. Mitra, G. Ductile deformation zones and mylonites: the mechanical processes involved in the deformation of crystalline basement rocks. *Am. J. Sci.* **278**, 1057–1084 (1978).
21. Atkinson, B. K. Subcritical crack propagation in rocks: theory, experimental results and applications. *J. Struct. Geol.* **4**, 41–56 (1982).
22. Paterson, M. S. in *Superplasticity in Metals, Ceramics, and Intermetallics* (eds Mayo, M. J., Kobayashi, M. & Wadsworth, J.) 303–312 (MPR Symp. Proc. 196, Materials Research Society, 1990).
23. Hiraga, T., Nagase, T. & Akizuki, M. The structure of grain boundaries in granite-origin ultramylonite studied by high-resolution electron microscopy. *Phys. Chem. Miner.* **26**, 617–623 (1999).
24. Shigematsu, N., Fujimoto, K., Ohtani, T. & Goto, K. Ductile fracture of fine-grained plagioclase in the brittle-plastic transition regime: implication for earthquake source nucleation. *Earth Planet. Sci. Lett.* **222**, 1007–1022 (2004).
25. Rybacki, E., Wirth, R. & Dresen, G. High-strain creep of feldspar rocks: Implications for cavitation and ductile failure in the lower crust. *Geophys. Res. Lett.* **35**, L04304 (2008).
26. Tullis, J., Yund, R. & Farver, J. Deformation-enhanced fluid distribution in feldspar aggregates and implications for ductile shear zones. *Geology* **24**, 63–66 (1996).
27. Wark, D. A. & Watson, E. B. Effect of grain size on the distribution and transport of deep-seated fluids and melts. *Geophys. Res. Lett.* **27**, 2029–2032 (2000).
28. Regenauer-Lieb, K., Yuen, D. A. & Fusseis, F. Landslides, ice quakes, earthquakes: a thermodynamic approach to surface instabilities. *Pure Appl. Geophys.* (in the press).
29. Etheridge, M. A., Wall, V. J. & Vernon, R. H. The role of the fluid phase during regional metamorphism and deformation. *J. Metamorph. Geol.* **1**, 205–226 (1983).
30. Chokshi, A. H. Cavity nucleation and growth in superplasticity. *Mater. Sci. Eng. A* **410–411**, 95–99 (2005).

Author Contributions F.F. did the field work, the FESEM analyses and part of the synchrotron experiment, interpreted the data and co-wrote the paper. K.R.-L. designed the study and co-wrote the paper. J.L. processed and analysed the tomographic data. R.M.H. designed and did part of the synchrotron experiment and co-wrote the paper. F.D.C. did part of the synchrotron experiment and processed tomographic data. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.F. (fusseis@cyllene.uwa.edu.au).

LETTERS

Colorado Plateau magmatism and uplift by warming of heterogeneous lithosphere

Mousumi Roy[1], Thomas H. Jordan[2] & Joel Pederson[3]

The forces that drove rock uplift of the low-relief, high-elevation, tectonically stable Colorado Plateau are the subject of long-standing debate[1–5]. While the adjacent Basin and Range province and Rio Grande rift province underwent Cenozoic shortening followed by extension[6], the plateau experienced ~2 km of rock uplift[7] without significant internal deformation[2–4]. Here we propose that warming of the thicker, more iron-depleted Colorado Plateau lithosphere[8–10] over 35–40 Myr following mid-Cenozoic removal of the Farallon plate from beneath North America[11,12] is the primary mechanism driving rock uplift. In our model, conductive re-equilibration not only explains the rock uplift of the plateau, but also provides a robust geodynamic interpretation of observed contrasts between the Colorado Plateau margins and the plateau interior. In particular, the model matches the encroachment of Cenozoic magmatism from the margins towards the plateau interior at rates of 3–6 km Myr^{-1} and is consistent with lower seismic velocities[13] and more negative Bouguer gravity[14] at the margins than in the plateau interior. We suggest that warming of heterogeneous lithosphere is a powerful mechanism for driving epeirogenic rock uplift of the Colorado Plateau and may be of general importance in plate-interior settings.

Rock and surface uplift in high-elevation, plate-interior regions, such as the Ordos block of the North China craton or the Colorado Plateau in the western United States, are difficult to ascribe to plate-boundary deformation processes. These regions generally lack deformation[2–4,15] and are underlain by depleted lithosphere[8–10,15], and the relationship between their surface and/or rock uplift and plate-tectonic forces is unclear.

Previous ideas for Colorado Plateau rock and/or surface uplift fall into four categories: early- to mid-Cenozoic Laramide-orogeny-related shortening[2,4,5]; mid- to late-Cenozoic epeirogeny[3]; stream incision[1,16,17] and isostatic responses[7]; and dynamic uplift[18]. Here we show that even if the contributions from minor Laramide deformation[4] and flexural isostatic responses to extension at the plateau margins and to net Cenozoic erosion are removed, there is >1.6 km of residual rock uplift that must be explained by post-Laramide tectonic processes. Dynamic uplift mechanisms can drive only 400–500 m of this residual amount[18], leaving ~1.2 km of unexplained rock uplift. We propose thermal perturbation and re-equilibration as a general mechanism for driving rock uplift within plate interiors, particularly in regions of thicker, more depleted lithosphere adjacent to zones of extension, such as the Colorado Plateau. Our model differs from previous ideas of thermal modification of the Colorado Plateau[3] in that it relies on a post-Laramide process that is triggered by the removal of the Farallon slab and the onset of thinning in the Basin and Range and Rio Grande rift provinces. We show that thermal perturbation following mid-Tertiary removal of the Farallon slab can account for the majority of the observed rock uplift of the

Colorado Plateau and, additionally, that this mechanism explains the observed rates of encroachment of the onset of Cenozoic magmatism onto the plateau.

The widespread distribution of shallow-marine and coastal-affinity sediments of late-Cretaceous age on the Colorado Plateau suggests relatively uniform elevation of the region at or near sea-level at that time. The present-day elevations of these strata can be reconstructed (including eustasy) to determine spatially varying net rock uplift and erosion functions across the plateau[7], with a mean of ~1.9 km net Cenozoic rock uplift (Fig. 1a) and net Cenozoic erosion of ~481 m (Fig. 1b). These reflect net deposition in early-Cenozoic time, mid-Cenozoic stability and then enhanced erosion since drainage integration in mid- to late-Cenozoic time[7].

Here we isolate the rock uplift of the Colorado Plateau that cannot be explained by Laramide crustal thickening, net Cenozoic erosion or flexural effects of extension in adjacent regions, and then model this residual rock uplift as being driven by thermal buoyancy modification. To remove contributions to the Cenozoic rock uplift of the Colorado Plateau from short-wavelength Laramide features (Fig. 1a), we smooth with a 2°-by-2° moving window (Fig. 1a, inset; Methods). In addition, we calculate the combined flexural response to spatially variable Cenozoic erosion (Fig. 1b) on the plateau and extension in adjacent regions (Methods). For reasonable flexural rigidities estimated for the Colorado Plateau[19], combined erosional and extensional unloading can drive ~0.5–0.6 km of rock uplift at the centre of the plateau and ~1 km at the edges that have undergone greater erosion or extension (Fig. 1c). Subtracting this response to unloading (Fig. 1c) from the smoothed rock uplift function (Fig. 1a, inset), we find the residual rock uplift (Fig. 1d). The average residual rock uplift increases from ~1.6 km at the centre of the plateau to a maximum of ~1.9–2.0 km approaching the margins, and then decreases at the extended western and southeastern margins (Fig. 1d). This residual cannot be explained by Laramide shortening or Cenozoic erosion and extension, and we model it here using mid-Tertiary buoyancy modification.

During Cenozoic time, the Colorado Plateau underwent subdued late-Mesozoic/early-Cenozoic shortening and minor extension, and was relatively unscathed by the voluminous and widespread mid-Tertiary magmatism (the 'ignimbrite flare-up'; Fig. 2) that affected regions surrounding the plateau[11,12,20]. Central to the idea of thermally driven rock uplift is transient heating of regions with thick lithosphere that protrude into the asthenosphere. Our hypothesis is that the Colorado Plateau represents one such region, where the difference in Cenozoic lithospheric thickness between the plateau and its surroundings[21] may be due to a combination of inherent chemical heterogeneity within North America[8–10], thinning of the plate in regions surrounding the plateau due to Cenozoic extension[6,21] and/or lithospheric modification in the Basin and Range

[1]Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA. [2]Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA. [3]Department of Geology, Utah State University, Logan, Utah 84322, USA.
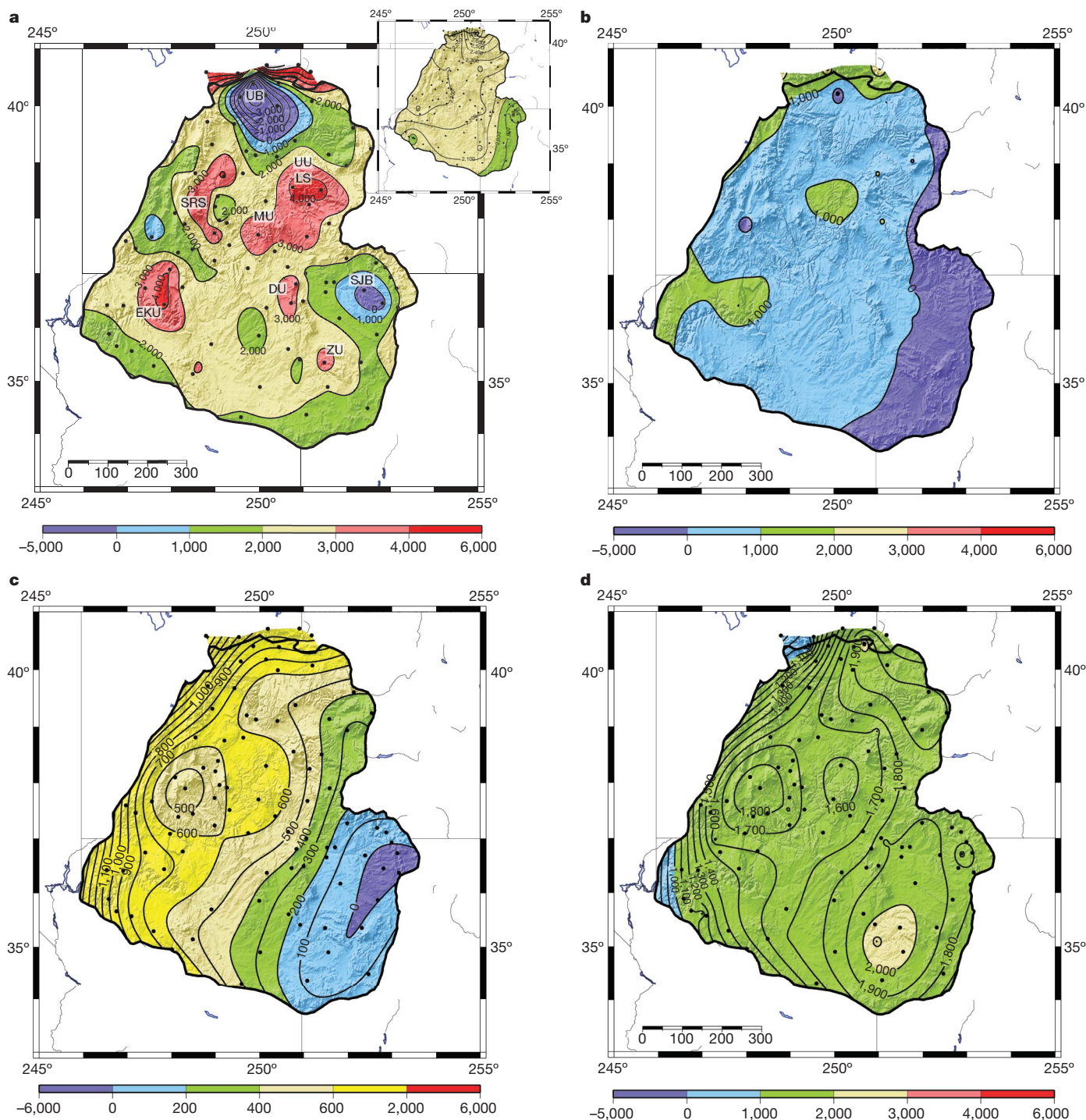
Figure 1 | Data (metres) used to derive the modelled residual rock uplift function. See Methods for details of model. **a**, Net Cenozoic (0–65 Myr ago) rock uplift across the Colorado Plateau[7], determined by stratigraphic constraints from field relations at the points marked (black dots). Laramide features: Uinta basin (UB), San Rafael swell (SRS), Uncompahgre uplift (UU), Monument uplift (MU), east Kaibab uplift (EKU), Defiance uplift (DU), San Juan basin (SJB), Zuni uplift (ZU), La Sal Mountains (LS). Inset, 2°-by-2° smoothed rock uplift function that effectively removes Laramide-related features. **b**, Net Cenozoic erosion function (smoothed rock uplift minus smoothed surface elevation; negative numbers denote burial). **c**, Flexural response to net Cenozoic erosion on the Colorado Plateau and extension in regions adjacent to the plateau (Methods), determined using a continuous plate with effective elastic plate thickness that was varied from $T_e = 20$ to 30 km (ref. 19). Representative results are shown for $T_e = 20$ km. **d**, Residual rock uplift function after flexural responses to erosion and extension (**c**) are removed from the smoothed rock uplift (inset in **a**). Scale bars in **a**, **b** and **d** show horizontal map distances in kilometres.

province upon removal of the Farallon slab roughly 30–40 Myr ago[11,12] (Fig. 3a).

Evidence for a chemically distinct Colorado Plateau lithosphere is seen in iron-depleted major-element compositions of mantle xenoliths from the western United States (Methods), suggesting that the plateau has undergone greater melt extraction than surrounding regions[8–10]. The enigmatic tectonic stability of the plateau maybe a result of this depletion[10], as the residuum following basalt removal is buoyant and/or of higher viscosity[22,23] and resists convective disruption[24] over timescales of 1 Gyr. Xenoliths also show that the Colorado Plateau may be in isopycnic equilibrium[10], with thicker lithosphere beneath the plateau than its surroundings. High seismic wave speeds
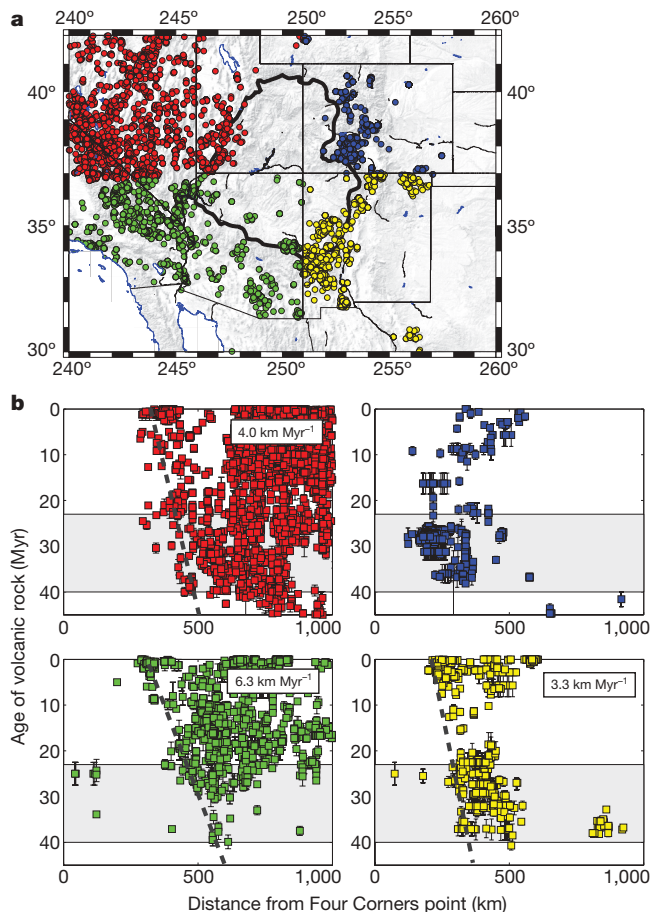
Figure 2 | Cenozoic magmatic patterns in the western United States, showing magmatic encroachment onto the Colorado Plateau.
a, Distribution of Cenozoic volcanic rocks in the Western North American Volcanic and Intrusive Rock Database with age uncertainties of less than 5 Myr, colour-coded by azimuth relative to the 'Four Corners' point, where Colorado, New Mexico, Arizona and Utah meet. Locations do not include structural reconstruction (for example in highly extended regions), but this is not a great source of error in undeformed regions such as the Colorado Plateau (outlined). **b**, Igneous rock age as a function of distance from the Four Corners point, colour-coded by quadrant as in **a**. The dashed lines illustrate the encroachment of the onset of magmatism onto the plateau at the margins adjacent to Neogene extensional regions (northwest, southeast and southwest) at the rates indicated. The grey shading indicates the time of the late-Eocene/Oligocene ignimbrite flare-up, inferred to be the time of removal of the Farallon slab[11,12]. By restricting attention to volcanic rocks, we avoid including cooling ages of intrusions, but the rates of magmatic encroachment observed here are unchanged if all igneous rocks are used instead. Error bars indicate the minimum and maximum reported ages for a particular sample.

exist beneath the plateau to depths of >100 km, whereas adjacent regions have slower wave speeds, suggesting thicker lithosphere beneath the plateau[12,13]. In the following, therefore, we assume that the Colorado Plateau lithosphere is thicker than its surroundings owing to a combination increased depletion of the plateau and thinning of adjacent regions, although we do not require the plateau to be isopycnic.

The Colorado Plateau is characterized by a lack of Cenozoic magmatism (Fig. 2a) and an encroachment of the onset of magmatism onto the plateau since mid-Tertiary time[25] (we define this encroachment as an enveloping surface on the space-time pattern of magmatism; Fig. 2b). The voluminous mid-Tertiary (23–40 Myr ago) ignimbrite flare-up accompanied the transition in deformation styles in the western United States from Mesozoic subduction to Neogene–present extension. It has been suggested that hydration of North America during Mesozoic subduction followed by the removal of
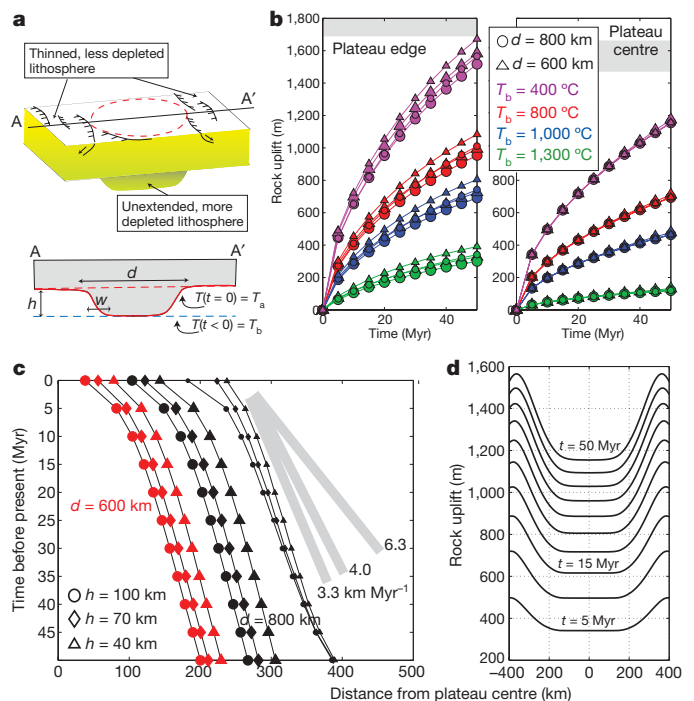

Figure 3 | Thermal and isostatic evolution during re-equilibration following a thermal perturbation at the base of a variable-thickness plate. a, Sketch of a heterogeneous plate, modelled as a cylindrically symmetric region that is thicker owing to a combination of greater depletion and lack of extension. **b**, Rock uplift at the edge and at the centre of the 'plateau' over 50 Myr of conductive relaxation for different values of d ($w = d/10$ and $h = 100$ km) and T_b, the initial basal temperature before mid-Tertiary time (modelled as $t = 0$), assumed to be the time of the thermal perturbation due to removal of the Farallon plate. The smaller symbols correspond to $T_e = 20$ km and the larger to $T_e = 30$ km (ref. 19). The observed range in residual rock uplift of the Colorado Plateau (Fig. 1d) is indicated by the grey regions. **c**, Lateral migration of the 1,200 °C isotherm at ~180-km depth as a function of d and h (results shown are for $T_b = 1,300$ °C). The slope of the lines represents the migration rate, which is a strong function of w, as illustrated by the lines with the smallest black symbols and fastest rate of encroachment: $d = 800$ km and $w = d/100$. All other results are shown for a larger value, $w = d/10$. Grey bars indicate rates observed in the migration of the onset of magmatism in Fig. 2b. **d**, Spatiotemporal pattern of rock uplift (every 5 Myr) for $d = 800$ km, $w = d/100$, $T_e = 30$ km and $T_b = 400$ °C, showing higher values at the margins than in the interior and migration of the point of highest uplift slightly in from the margins of the plateau.

the Farallon plate drove the voluminous magmatic response[11,12,20], and we assume that mid-Tertiary time coincided with the transition from cooler mantle-wedge conditions beneath the plate[26,27] to hotter conditions as hot asthenosphere contacted the base of the plate. We propose that transient heating of the thicker Colorado Plateau lithosphere was triggered by the mid-Tertiary thermal perturbation and was aided by the fact that regions surrounding the plateau experienced coeval plate thinning. In our models, the contrast in thermal boundary thickness between the Colorado Plateau and its surroundings evolves as the plateau warms, but we ignore time-dependent extension of surrounding regions.

We combine flexure with three-dimensional conductive models of heat transfer (Methods Summary) over a 40–50-Myr interval in an idealized variable-thickness plate representing the combined effects of greater depletion in the central part and extensional thinning of the surrounding, less depleted regions (Fig. 3a; Methods). The 50-Myr model time interval is chosen to encompass the time since the ignimbrite flare-up in the western United States (~25–40 Myr ago) and captures expected thermal effects in the Colorado Plateau following removal of the Farallon slab[11,12].

Our models indicate that transient warming since mid-Tertiary time (defined as $t = 0$) thins the thermal boundary layer and drives

rock uplift of the Colorado Plateau over the last 20–40 Myr (Fig. 3b), producing greater rock uplift at the margins than in the plateau interior, with the maximum rock uplift occurring just inside the plateau margins (Fig. 1d). Buoyancy modification is driven by the inward migration of a conductive heating front (represented, for example, by the lateral migration of the $T = 1,200\,°C$ isotherm measured at a depth of ~ 180 km) at a rate that is governed by the length scale, w, over which the plate thickness changes (Fig. 3a) and has a minor dependence on the excess thickness, h (Fig. 3c).

The expected amplitude of rock uplift from conductive heating depends on the initial thermal structure and the shape of the plate (Fig. 3b; Methods). For a diameter of $d = 800$ km, comparable to that of the Colorado Plateau, and an initial basal temperature of $T_b = 400\,°C$, we expect rock uplift of 1.2–1.4 km at the centre and 1.6–1.7 km at the edges in 30 Myr. This low initial basal temperature corresponds to a plate that is insulated by the presence of the Farallon slab (before thermal relaxation), consistent with xenolith palaeothermometry and thermal modelling of the Laramide orogeny[26,27]. Although shallow subduction of the Farallon plate has been inferred from thermal histories of rocks[26,27], we note that our models do not explicitly require it. Our models require a mid-Tertiary thermal perturbation at the base of the Colorado Plateau and surroundings, and our results lend support to the idea that removal of a shallowly subducting Farallon plate would provide such a thermal perturbation.

The predicted timing and spatial distribution of rock uplift may be tested using detailed studies of rock cooling and exhumation together with stratigraphic constraints[16,17]. We note that the southern margin of the plateau has experienced greater early-Cenozoic exhumation but little late-Cenozoic erosion outside the narrow Grand Canyon[7,17]. In contrast, the plateau interior has been affected by a pulse of late-Cenozoic incision along the trunk drainage system[7], decoupled from the timing of thermally driven rock uplift (Fig. 1b). A regionally extensive data set of long-term exhumation is required to elucidate this further and separate the progressive uplift predicted in our model (Fig. 3b) from the signal of more recent base-level fall along the drainage system, although periods of rapid exhumation are easier to detect in rock cooling data than are slow, protracted rock uplift and erosion.

Although we do not explicitly include partial melting in our model, if we use the $T = 1,200\,°C$ isotherm as a first-order proxy for the onset of hydrous melting in the upper mantle[28], the rates of inward migration of the heating front are predicted to be 3–6 km Myr^{-1} for a wide range of plateau diameters ($d = 600$–800 km, comparable to the width of the Colorado Plateau), consistent with the inferred rates at which the onset of magmatism encroaches onto the plateau (Figs 2 and 3c). We also observe a late-stage increase in this rate as vertically rising and inward-migrating conductive heating fronts pass a given point (change in slope between 0 and 5 Myr in Fig. 3c). Future, more detailed, comparisons with phase relationships in a melting model must incorporate variable chemistry and hydration of source regions[12] and changes in both chemical and thermal buoyancy during and following the mid-Tertiary ignimbrite flare-up[14].

The observed asymmetry in the rate of encroachment of the onset of Cenozoic magmatism onto the Colorado Plateau (4–6.3 km Myr^{-1} at the northwestern and southwestern margins adjacent to the highly extended Basin and Range province, versus 3.3 km Myr^{-1} at the southeastern margin adjacent to the significantly less extended Rio Grande rift province) is predicted by our model to be a consequence of variations in the excess thickness, h, of the plateau relative to its surroundings and the horizontal scale, w, over which plate thickness varies (w is predicted to be shorter at the margins adjacent to the Basin and Range province than at the margin adjacent to the Rio Grande rift province). Our models predict the upper mantle at the margins of the Colorado Plateau to be warmer and lower in density than the upper mantle in its interior, consistent with lower upper-mantle seismic velocities[12,13] and lower Bouguer gravity[14]. Additionally, the magmatic alteration of the margins of the

plateau predicted by our thermal model is consistent with xenolith populations that at the margins of the plateau contain more metasomatized, pyroxenite-bearing rocks than they do in the interior[14]. In particular, the presence of garnet-websterite xenoliths hosted in 3-Myr-old volcanic fields at the margin of the Colorado Plateau[14] is consistent with metasomatic alteration[9,14,26] at the edges of the intact but altered plateau lithosphere. Finally, comparisons of inferred upper-mantle temperatures from seismic velocity and heat flow show evidence for a transient thermal regime in the plateau and the Rio Grande rift[29], at the margins of the plateau.

To conclude, the predicted pattern of thermally driven mid-Tertiary–present rock uplift of the Colorado Plateau, with smaller rock uplift in the interior than at the margins (Fig. 3b, d), is comparable to the patterns observed in the residual rock uplift of the plateau (Fig. 1d). Conductive re-equilibration explains a large fraction of the residual rock uplift, although to match the full magnitude of the rock uplift (particularly in the central plateau) and to match local highs and lows, models will probably need to include the effects of variable plate rigidity, spatially variable extension and dynamic uplift[18]. Additionally, part of the discrepancy between predicted and observed rock uplift may be due to non-zero average Laramide rock uplift. Our model nevertheless provides a coherent thermal process that explains disparate observations such as the pattern and rate of encroachment of the onset of Cenozoic magmatism from the margins to the plateau interior (Figs 2b and 3c), lower seismic velocities[12,13] at the margins than in the plateau interior and lower-density upper mantle at the margins than in the interior (as inferred from xenoliths and gravity[14]), and patterns of rock cooling across the southern plateau[17]. Thermal perturbation, re-equilibration and the accompanying magmatism in plate-interior settings is likely to be an important process for the rejuvenation and refertilization of cratonic regions and may drive significant intra-plate epeirogenic rock uplift.

METHODS SUMMARY

We calculate flexural isostatic responses by convolving the expected local response with the two-dimensional Green's function for an elastic plate of uniform rigidity[30]. We use two rigidities, corresponding to effective elastic plate thicknesses of $T_e = 20$ and 30 km, chosen to bracket the value of $T_e = 22$–25 km estimated from coherence of gravity and topography for the Colorado Plateau[19]. The response of the Colorado Plateau to extensional unloading at margins adjacent to the Basin and Range and Rio Grande rift provinces is calculated using published basin loads from joint inversions of gravity and topography (Methods).

We assume a linear initial geotherm with $T_0 = 0\,°C$ at the surface and a basal temperature of $T_b = 400, 800, 1,000$ or $1,300\,°C$ (Fig. 3b). These T_b values represent normal to cooler-than-normal mantle-wedge conditions[26,27] before the removal of the Farallon slab at $t = 0$. Immediately after $t = 0$, the plate is subjected to a new basal temperature, $T_a = 1,300\,°C$ and allowed to re-equilibrate. We assume a thermal diffusivity of $\kappa = 1 \times 10^{-6}\,m^2\,s^{-1}$, zero upper-mantle heat production and uniform crustal heat production of $3 \times 10^{-6}\,W\,m^{-3}$ over a depth range of 0–15 km. Conduction drives lateral and basal heating of the thicker lithosphere, so initially warped isotherms flatten out. The assumption of uniform temperatures beneath the plate (ignoring the adiabatic gradient) is equivalent to the assumption that convection maintains contact between hot asthenosphere and the plate.

We use the time-dependent thermal structure to derive a time-dependent density (assuming a coefficient of thermal expansion of $\alpha = 2.5 \times 10^{-5}\,K^{-1}$) and transient flexural isostatic responses. The conductive timescale for the steady state is ~ 1 Gyr, but warming of the Colorado Plateau is enhanced by three-dimensional conduction inward from the margins of the ~ 600–800-km-diameter region. The conductive timescale, τ, scales with area ($\tau \approx (\text{area})^{1/2}/\kappa$) and our models show that for the typical diffusivity assumed, transient responses over ~ 10–100 Myr will be a significant fraction of the total steady-state response for regions of length scale 10^2 km.

1. Hunt, C. B. Cenozoic geology of the Colorado Plateau. *Prof. Pap. US Geol. Surv.* **279** (1956).

2. Bird, P. Continental delamination and the Colorado Plateau. *J. Geophys. Res.* **84**, 7561–7571 (1979).
3. Morgan, P. & Swanberg, C. A. On the Cenozoic uplift and tectonic stability of the Colorado Plateau. *J. Geodyn.* **3**, 39–63 (1985).
4. Spencer, J. E. Uplift of the Colorado Plateau due to lithospheric attenuation during Laramide low-angle subduction. *J. Geophys. Res.* **101**, 13595–13609 (1996).
5. McQuarrie, N. & Chase, C. G. Raising the Colorado Plateau. *Geology* **28**, 91–94 (2000).
6. Wernicke, B., Christiansen, R. L., England, P. C. & Sonder, L. J. Tectonomagmatic evolution of Cenozoic extension in the North American Cordillera. *Spec. Publ. Geol. Soc. (Lond.)* **28**, 203–221 (1987).
7. Pederson, J. L., Mackley, R. D. & Eddleman, J. L. Colorado Plateau uplift and erosion evaluated using GIS. *GSA Today* **12**, 4–10 (2002).
8. Alibert, C. Peridotite xenoliths from the western Grand Canyon and the Thumb: a probe into the subcontinental mantle of the Colorado Plateau. *J. Geophys. Res.* **99**, 21605–21620 (1990).
9. Smith, D. Insights into the evolution of the uppermost continental mantle from xenolith localities on and near the Colorado Plateau and regional comparisons. *J. Geophys. Res.* **105**, 16769–16781 (2000).
10. Lee, C.-T. *et al.* Preservation of ancient and fertile lithospheric mantle beneath the southwestern United States. *Nature* **411**, 69–73 (2001).
11. Humphreys, E. D. Post-Laramide removal of the Farallon slab, western United States. *Geology* **23**, 987–990 (1995).
12. Humphreys, E. D. *et al.* How Laramide-age hydration of North American lithosphere by the Farallon slab controlled subsequent activity in the western United States. *Int. Geol. Rev.* **45**, 575–595 (2003).
13. Sine, C. R. *et al.* Mantle structure beneath the western edge of the Colorado Plateau. *Geophys. Res. Lett.* **35**, doi:10.1029/2008GL033391 (2008).
14. Roy, M., MacCarthy, J. K. & Selverstone, J. Upper mantle structure beneath the eastern Colorado Plateau and Rio Grande rift revealed by Bouguer gravity, seismic velocities, and xenolith data. *Geochem. Geophys. Geosyst.* **6**, doi:10.1029/2005GC001008 (2005).
15. Menzies, M., Xu, Y., Zhang, H. & Fan, W. Integration of geology, geophysics and geochemistry: a key to understanding the North China Craton. *Lithos* **96**, 1–21 (2007).
16. McMillan, M. E., Heller, P. L. & Wing, S. L. History and causes of post-Laramide relief in the Rocky Mountain orogenic plateau. *Bull. Geol. Soc. Am.* **118**, 393–405 (2006).
17. Flowers, R., Wernicke, B. P. & Farley, K. A. Unroofing, incision and uplift history of the southwestern Colorado Plateau from (U-Th)/He apatite thermochronometry. *Bull. Geol. Soc. Am.* **120**, 571–587 (2008).
18. Moucha, R. *et al.* Mantle convection and the recent evolution of the Colorado Plateau and the Rio Grande Rift valley. *Geology* **36**, 439–442 (2008).
19. Lowry, A. & Smith, R. B. Flexural rigidity of the Basin and Range-Colorado Plateau-Rocky Mountain transition from coherence analysis of gravity and topography. *J. Geophys. Res.* **99**, 20123–20140 (1994).
20. Lipman, P. W. & Glazner, A. F. Introduction to middle Tertiary Cordilleran volcanism – Magma sources and relations to regional tectonics. *J. Geophys. Res.* **96**, 13193–13199 (1991).
21. Wang, K., Plank, T., Walker, J. D. & Smith, E. I. A melting profile across the Basin and Range, SW USA. *J. Geophys. Res.* **107**, doi:10.1029/2001JB000209 (2002).
22. Kelly, R. K., Kelemen, P. B. & Jull, M. Buoyancy of the continental upper mantle. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2002GC000399 (2003).
23. Schutt, D. L. & Lesher, C. E. Effects of melt depletion on the density and seismic velocity of garnet and spinel lherzolite. *J. Geophys. Res.* **111**, doi:10.1029/2003JB002950 (2006).
24. Lenardic, A., Moresi, L.-N. & Mülhaus, H. Longevity and stability of cratonic lithosphere: insights from numerical simulations of coupled mantle convection and continental tectonics. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001859 (2003).
25. Wenrich, K. J. *et al.* Spatial migration and compositional changes of Miocene-Quaternary magmatism in the western Grand Canyon. *J. Geophys. Res.* **100**, 10417–10440 (1995).
26. Smith, D. & Griffin, W. L. Garnetite xenoliths and mantle-water interactions below the Colorado Plateau, southwestern United States. *J. Petrol.* **46**, 1901–1924 (2005).
27. English, J. M., Johnston, S. T. & Wang, K. L. Thermal modeling of the Laramide orogeny: testing the flat-slab subduction hypothesis. *Earth Planet. Sci. Lett.* **214**, 619–632 (2003).
28. Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parametrization of hydrous melting. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2002GC000433 (2003).
29. Goes, S. & van der Lee, S. Thermal structure of the North American uppermost mantle inferred from seismic tomography. *J. Geophys. Res.* **107**, doi:10.1029/2000JB000049 (2002).
30. Jones, C. H., Unruh, J. R. & Sonder, L. J. The role of gravitational potential energy in active deformation in the southwestern United States. *Nature* **381**, 37–41 (1996).

nature

METHODS

Data analysis. The estimates of net Cenozoic rock uplift across the Colorado Plateau (Fig. 1a) are based on outcrops and reconstructions of the base of late-Cretaceous marine-affinity strata[7], including a eustatic correction for higher Cretaceous sea level (Supplementary Information). At the majority of locations where the base of the late-Cretaceous marine-affinity strata outcrops, uncertainties from estimating the present-day elevation of this stratigraphic datum are <1 m; at the minority of locations where the strata are reconstructed, uncertainties are larger, but are generally <100 m. Additional minor uncertainties (<1 m) arise from uncertainties in the 90-m digital elevation model used[7].

We note that Laramide features correspond to locally high rock uplift. To remove these short-wavelength features and focus on regional patterns, we smooth the rock uplift function with a moving-window average (the results are unchanged for windows ranging from 1° by 1° to 2° by 2°; Fig. 1a, inset). The northern margin of the plateau, which has not undergone extension (in contrast to the western, southern and eastern margins), has large positive rock uplift values (>5 km) in the unextended, Laramide-age Uinta Mountains; after smoothing, these off-plateau points lead to positive rock uplift numbers in the Uinta basin (Fig. 1a, inset). The smoothed net Cenozoic erosion function (Fig. 1b) is the difference between the smoothed net Cenozoic rock uplift and smoothed present-day surface elevation (the topography was smoothed by averaging using the same windows as the rock uplift function).

Flexural response of the Colorado Plateau to adjacent extension. The regions to the west, south and east of the plateau have undergone variable extension over the past 30 Myr, and the flexural response to this extension should affect rock uplift at the margins of the plateau. The flexural response at the margins of the highly extended Basin and Range province is difficult to estimate because of the wide, distributed nature of the deformation complicated by magmatic loads[6], whereas the more compact, relatively amagmatic basins of the Rio Grande rift allow joint analysis of gravity and topography to estimate basin loads[31]. Upward-directed basin loads estimated in the Rio Grande rift range from $3.5 \times 10^{11}\,\mathrm{N\,m^{-1}}$ (northern Albuquerque basin, 17–28% extension) to $1.6 \times 10^{12}\,\mathrm{N\,m^{-1}}$ (Tularosa basin, 40–50% extension)[31]. We use the higher value in our calculations of the flexural response to extension at the margins of the Colorado Plateau and assume that the load is imposed uniformly over all regions outside the plateau. The flexural response to this load within the plateau (Supplementary Fig. 1) is then added to the flexural response to net Cenozoic erosion[7], and this sum is subtracted from the smoothed rock uplift function (Fig. 1a, inset).

The extensional loads assumed here are idealized, and we acknowledge that extension surrounding the plateau is spatially variable. This model provides a first-order estimate of the effects of extension surrounding the western, southern and eastern margins of the plateau. However, the northern and northeastern margins are not extended, so the calculated flexure is not relevant in this area and we caution that the resulting inferred residual rock uplift in the Uinta basin (Fig. 1d) is probably an underestimate. By assuming that large-magnitude extension uniformly surrounds the Colorado Plateau, however, we obtain a maximum estimate for rock uplift due to spatially variable extension adjacent to the plateau. This is an appropriate first-order approach because we wish to determine the minimum residual rock uplift that is unexplained by processes such as erosion and extension. Our assumption of a uniform-rigidity plate is likely to be invalid at the margins of the extensional zones adjacent to the plateau, which are probably of lower rigidity[19,31] (alternatively, the plate may not be continuous), so the residual rock uplift magnitude is probably an overestimate. In this work, we consider the effects of net Cenozoic erosion (Fig. 1b) and net Cenozoic extension (Supplementary Fig. 1), ignoring any temporal variations in the onset and duration of extension adjacent to the Colorado Plateau.

Magnesium number of on- and off-plateau xenoliths. We use published xenolith compositions[8–10,32–56] and divide the rocks into on- and off-plateau groups (Supplementary Fig. 2). Xenoliths on the plateau are distinctly more magnesian than those off the plateau, both for whole-rock[32–40] and olivine[8,32–56] analyses (Supplementary Fig. 2).

Thermal model. In our model, the initial thermal state corresponds to flat isotherms ($t < 0$; Fig. 3a), but at $t = 0$ we impose a thermal perturbation such that the base of the variable-thickness plate is an isotherm at $T_a = 1{,}300\,^{\circ}\mathrm{C}$. The thicker plate in the Colorado Plateau is assumed to reflect a combination of greater depletion and a lack of Cenozoic extension compared with the surroundings. We solve the heat equation for $t > 0$ within the variable-thickness thermal boundary layer by assuming that the surface and basal temperatures are fixed, at $T_0 = 0\,^{\circ}\mathrm{C}$ and $T_a = 1{,}300\,^{\circ}\mathrm{C}$, respectively. The initial linear temperature profile, $T(z) = (T_b - T_0)/h(x)$, where z is depth and x is horizontal position, means that even if the assumed initial basal temperature is $T_b = 1{,}300\,^{\circ}\mathrm{C}$, there will be some thermal evolution in the model owing to both three-dimensional effects and assumed crustal heat production (Supplementary Fig. 3), which leads to a minor contribution to the total inferred rock uplift (Fig. 3b).

As temperatures equilibrate in the thermal boundary layer, the rock uplift calculated in the model is consistent with rapid rates of rock uplift immediately upon imposition of the basal perturbation, with higher uplift rates at the margins of the plateau than in the interior (Fig. 3d). Uplift rates decrease through time as the system approaches the steady state, and the final rock uplift is determined by both the final thermal structure and the assumed shape of the thermal boundary layer. In the steady state, the isotherms will be smoother than they were initially, but will mimic the shape of the thermal boundary layer (in the same manner that isotherms mimic topography near the surface). We assume that the transition from the thicker thermal boundary layer (200 km) beneath the plateau to the thinner thermal boundary layer (200 km − h; Fig. 3a) in the surrounding regions is given by the function $\mathrm{atan}(x/w)$, with a length scale of $w = d/10$ or $d/100$. The steady-state rock uplift roughly mirrors the shape of the protrusion in the thermal boundary layer, so although the rock uplift starts at a higher rate at the margins (Fig. 3b) and uplift rates slow down through time everywhere, the central region is predicted to continue uplifting longer than the margins. As a result, the point with maximum rock uplift at a given time migrates inward from the margin to the plateau interior (Fig. 3d).

31. Peterson, C. & Roy, M. in *Geology of the Chama Basin* (eds Brister, B. S., Bauer, P. W. & Read, A. S) 105–114 (56th Field Conf. Guidebook, New Mexico Geological Society, 2005).
32. Berman, S. C., Foland, K. A. & Spera, F. J. On the origin of an amphibole-rich vein in a peridotite inclusions from the Lunar Crater Volcanic Field, Nevada, U.S.A. *Earth Planet. Sci. Lett.* **56,** 343–361 (1981).
33. Condie, K. C., Cox, J., O'Reilly, S. Y., Griffin, W. L. & Kerrich, R. Distribution of high field strength and rare earth elements in mantle and lower crustal xenoliths from the Southwestern United States: the role of grain-boundary phases. *Geochim. Cosmochim. Acta* **68,** 3919–3942 (2004).
34. Ehrenberg, S. N. Petrogenesis of garnet lherzolite and megacrystalline nodules from the Thumb, Navajo Volcanic Field. *J. Petrol.* **23,** 507–547 (1982).
35. Feigenson, M. D. Continental alkali basalts of kimberlite and depleted mantle: evidence from Kilbourne Hole maar, New Mexico. *Geophys. Res. Lett.* **13,** 965–968 (1986).
36. Lauglin, A. W. & Brookins, D. G. Chemical and strontium isotopic investigations of ultramafic inclusions and basalt, Bandera Crater, New Mexico. *Geochim. Cosmochim. Acta* **35,** 107–113 (1971).
37. McGuire, A. V. & Mukasa, S. B. Magmatic modification of the uppermost mantle beneath the Basin and Range to Colorado Plateau Transition Zone; evidence from xenoliths, Wikieup, Arizona. *Contrib. Mineral. Petrol.* **128,** 52–65 (1997).
38. Roden, M. F., Irving, A. J. & Murthy, V. R. Isotopic and trace element composition of the upper mantle beneath a young continental rift: results from Kilbourne Hole, New Mexico. *Geochim. Cosmochim. Acta* **52,** 461–473 (1988).
39. Smith, D. & Riter, J. C. A. Genesis and evolution of low-Al orthopyroxene in spinel peridotite xenoliths, Grand Canyon field, Arizona, USA. *Contrib. Mineral. Petrol.* **127,** 391–404 (1997).
40. Wilshire, H. G. *et al.* Mafic and ultramafic xenoliths from volcanic rocks of the western United States. *Prof. Pap. US Geol. Surv.* **1443** (1988).
41. Baldridge, W. S. Mafic and ultramafic inclusion suites from the Rio Grande rift (New Mexico) and their bearing on the composition and thermal state of the lithosphere. *J. Volcanol. Geotherm. Res.* **6,** 319–351 (1979).
42. Beard, B. L. & Glazner, A. F. Trace element and Sr and Nd isotopic composition of mantle xenoliths from the Big Pine volcanic field, California. *J. Geophys. Res.* **100,** 4169–4181 (1985).
43. Bussod, G. Y. A. *Thermal and Kinematic History of Mantle Xenoliths from Kilbourne Hole, New Mexico.* Report No. LA-9616-T (Los Alamos National Laboratory, 1983).
44. Evans, S. H. Jr & Nash, W. P. Petrogenesis of xenolith-bearing basalts from southeastern Arizona. *Am. Mineral.* **64,** 249–267 (1979).
45. Fodor, R. V. Ultramafic and mafic inclusions and megacrysts in Pliocene basalt, Black Range, New Mexico. *Bull. Geol. Soc. Am.* **89,** 451–459 (1978).
46. Frey, F. A. & Prinz, M. Ultramafic inclusions from San Carlos, Arizona: petrologic and geochemical data bearing on their petrogenesis. *Earth Planet. Sci. Lett.* **38,** 129–176 (1978).
47. Galer, S. J. G. & O'Nions, R. K. Chemical and isotopic studies of ultramafic inclusions from the San Carlos Volcanic Field, Arizona: a bearing on their petrogenesis. *J. Petrol.* **30,** 1033–1064 (1989).
48. Hervig, R. L., Smith, J. V. & Dawson, J. B. Lherzolite xenoliths in kimberlites and basalts: petrogenetic and crystallochemical significance of some minor and trace elements in olivine, pyroxenes, garnet and spinel. *Trans. R. Soc. Edinb. Earth Sci.* **77,** 181–202 (1986).
49. Johnson, K. E. *An Appraisal of Mantle Metasomatism based upon Oxidation States, Trace Element and Isotope Geochemistry, and Fluid/Rock Ratios in Spinel Lherzolite Xenoliths.* PhD thesis, Northwestern Univ. (1990).
50. Riter, J. C. A. *Geochemical and Tectonic Evolution of the Colorado Plateau Mantle Lithosphere: Evidence from Grand Canyon Mantle Xenoliths.* PhD thesis, Univ. Texas, Austin (1999).
51. Rosenbaum, J. M., Zindler, A. & Rubenstone, J. L. Mantle fluids: evidence from fluid inclusions. *Geochim. Cosmochim. Acta* **60,** 3229–3252 (1996).
52. Smith, D. & Riter, J. C. A. Genesis and evolution of low-Al orthopyroxene in spinel peridotite xenoliths, Grand Canyon field, Arizona, USA. *Contrib. Mineral. Petrol.* **127,** 391–404 (1997).

53. Smith, D. & Levy, S. Petrology of Green Knobs diatreme, New Mexico, and implications for the mantle below the Colorado Plateau. *Earth Planet. Sci. Lett.* **19**, 107–125 (1976).

54. Roden, M. F. & Shimizu, N. Ion microprobe analyses bearing on the composition of the upper mantle beneath the Basin and Range and Colorado Plateau provinces. *J. Geophys. Res.* **98**, 14091–14108 (1993).

55. Tyner, G. N. & Smith, D. Peridotite xenoliths in silica-rich, potassic latite from the transition zone of the Colorado Plateau in north-central Arizona. *Contrib. Mineral. Petrol.* **94**, 63–71 (1986).

56. Wilshire, H. G., McGuire, A. V., Noller, J. S. & Turrin, B. D. Petrology of lower crustal and upper mantle xenoliths from the Cima Volcanic Field, California. *J. Petrol.* **32**, 169–200 (1991).

LETTERS

Adaptation and the evolution of parasite virulence in a connected world

Geoff Wild[1], Andy Gardner[2] & Stuart A. West[3]

Adaptation is conventionally regarded as occurring at the level of the individual organism, where it functions to maximize the individual's inclusive fitness[1–3]. However, it has recently been argued that empirical studies on the evolution of parasite virulence in spatial populations show otherwise[4–7]. In particular, it has been claimed that the evolution of lower virulence in response to limited parasite dispersal[8,9] provides proof of Wynne-Edwards's[10] idea of adaptation at the group level. Although previous theoretical work has shown that limited dispersal can favour lower virulence, it has not clarified why, with five different suggestions having been given[6,8,11–15]. Here we show that the effect of dispersal on parasite virulence can be understood entirely within the framework of inclusive fitness theory. Limited parasite dispersal favours lower parasite growth rates and, hence, reduced virulence because it (1) decreases the direct benefit of producing offspring (dispersers are worth more than non-dispersers, because they can go to patches with no or fewer parasites), and (2) increases the competition for hosts experienced by both the focal individual ('self-shading') and their relatives ('kin shading'). This demonstrates that reduced virulence can be understood as an individual-level adaptation by the parasite to maximize its inclusive fitness, and clarifies the links with virulence theory more generally[16].

Darwin's theory of evolution by natural selection explains both the process and the purpose of adaptation[17,18]. The process of adaptation occurs through the action of natural selection, which is mediated by differential reproductive success of individual organisms, and resulting changes in gene frequency[17]. This process leads individual organisms to appear designed as if for the purpose of maximizing their inclusive fitness, which is defined as the effect of one individual's actions on its genetic contribution to future generations through its direct descendants and those of its relatives[1,2]. The inclusive fitness approach to adaptation has been extremely successful, especially in the fields of behavioural and evolutionary ecology, providing explanations for a wide range of traits[19,20].

Despite the success of inclusive fitness theory, a number of recent papers have challenged the idea, arguing that natural selection can favour group adaptations in cases in which inclusive fitness is not maximized[4–7]. This suggestion is analogous to Wynne-Edwards's original idea of group selection[10], whereby adaptations occur for the benefit of the group. The primary empirical evidence upon which this challenge is based[4–7] is the experimental observation that parasites (viruses) of both moths and bacteria evolve to cause less damage to their hosts (lower virulence) in spatially structured populations, where dispersal can be limited[8,9]. The argument here is that the parasites become more prudent to prevent overexploitation and, hence, avoid causing the extinction of the local host population. However, it seems plausible that this effect of limited dispersal could also be explained by inclusive fitness theory, because it will lead to a higher relatedness between interacting parasites, which has long been

known to favour a more prudent exploitation of host resources and, therefore, a lower virulence[16] (Supplementary Information). The only way to resolve this debate is to move away from verbal arguments and towards formal theoretical models that incorporate explicit spatial dynamics such as variable patch sizes and within-patch demography, and to use such models to determine the underlying evolutionary mechanisms[21].

Here we address this problem by using a standard epidemiological model[16,22], in the context of a geographically structured population, to determine why limited parasite dispersal selects for lower levels of virulence. We assume the simplest possible situation to make the underlying selective forces explicit and to allow comparison with previous models, which have shown that dispersal influences virulence but have failed to clarify why[11–15]. More general discussion of the various ways in which virulence theory has been expanded, to examine the consequences of a range of potentially important biological factors, are provided elsewhere[16,23]. In addition to its role in the debate over the process of adaptation, this effect of dispersal may be particularly important for the evolution of parasites, because it suggests that as human activity makes the world more connected, natural selection will favour more virulent and dangerous parasites[12].

We assume an 'island model' with an infinite number of patches (subpopulations), each of which may contain up to N host individuals. In this model, an individual (host or parasite) either remains on its natal patch or disperses. If it disperses, each of the other patches in the population is an equally likely destination. The island model is a standard tool for examining the effect of population structure while allowing analytical simplifications by dividing interactions into those that are 'local' (same patch) and those that are 'global' (different patch)[21,24]. We assume that hosts reproduce at a constant per-capita rate, b. A newborn host will attempt to settle either on its natal patch (local dispersal), with probability $1 - d_h$, or on a randomly chosen, non-natal patch (global dispersal), with probability d_h. A newborn host successfully settles on a patch only when the patch in question supports fewer than N individuals. If successful, the newborn host is assumed, for convenience, to mature instantaneously. If unsuccessful, the newborn dies. We assume also that adult hosts are not capable of dispersal—each adult remains on the patch it settled as a newborn.

We classify hosts as either infected by the parasite or uninfected. We ignore the possibility of multiple infections[16], so infected and uninfected hosts might also be called non-susceptible and susceptible, respectively. In our model, uninfected hosts die at constant per-capita rate, μ. Infected hosts, on the other hand, suffer a greater risk of mortality, dying at rate $\mu + z$. Here z describes the disease-induced mortality (parasite virulence) that arises as a consequence of the parasite's exploitation of its host.

We assume that parasite transmissibility, $\beta(z)$, is positively correlated with parasite virulence (z), to reflect the standard assumption

[1]Department of Applied Mathematics, University of Western Ontario, London, Ontario N6A 5B7, Canada. [2]Institute of Evolutionary Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK. [3]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.
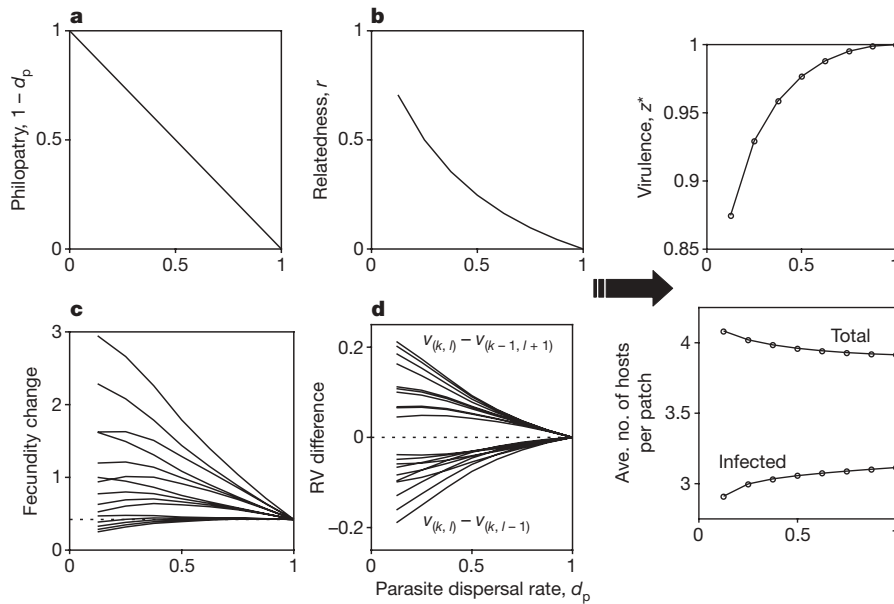
Figure 1 | Increasing parasite dispersal affects equation (1) in different ways. Increasing d_p decreases parasite philopatry (**a**); decreases the relatedness between parasite neighbours ($r_{(i,j)} = r$ for all i and $j > 2$; see Supplementary Information) (**b**); reduces the variation in parasite fecundity change across patch types (**c**); and reduces variation in reproductive value (RV) across parasite classes (**d**). The net effect of changes **a–d** is illustrated on the far right. Increasing d_p increases both the evolutionarily stable virulence level, z^*, and the fraction of hosts infected by the parasite. Results were generated by numerical simulation with $\mu = 1$, $d_h = 0.9$, $b = 3$, $N = 5$ and $\beta(x) = 5x/(1 + x)$.

that increased parasite growth leads not only to greater transmission, but also to greater host mortality[16,22]. We allow only horizontal transmission of the parasite, from infected adult to uninfected adult; hence, vertical transmission from host parent to its unborn offspring is not possible. Transmission is assumed to occur locally (within each patch), at a rate proportional to $(1-d_p)\beta(z)$, and globally (to other patches), at a rate proportional to $d_p\beta(z)$. In both cases, parasite transmission is assumed to follow a law of mass action. The parameter d_p is a proportion and is interpreted as the rate at which parasite offspring 'disperse' to new, randomly chosen patches.

We classify patches, and the parasites on those patches, according to the local number of uninfected (i) and infected (j) hosts. Naturally, parasite fitness depends upon the class to which its patch belongs, and the distribution of the different classes of patch in the population. To determine the evolutionarily stable level of parasite virulence, z^*, we consider a rare mutant parasite (the focal individual) belonging to class (k, l)—that is, on a patch with k uninfected hosts and l infected hosts. We note that although the global frequency of the mutant parasite is negligible, the probability that a parasite neighbour of a mutant is itself a mutant is not necessarily negligible. Thus, it is reasonable to expect that the social effects of mutant virulence are felt by other mutants as well.

In Supplementary Information, we show that if the focal mutant parasite increases its virulence phenotype by a small amount, $\delta > 0$, the resulting change in its inclusive fitness, $\Delta W_{(k, l)}$, is given by

$$
\begin{aligned}
\Delta W_{(k, l)} = {}& -\delta v_{(k, l)} \\
& + \delta\beta'(z)\left[(1-d_p)kv_{(k-1, l+1)} + d_p\sum\nolimits_{(i, j)} v_{(i, j)}(i+1)p_{(i+1, j-1)}\right] \\
& - \delta\beta'(z)(1-d_p)k(v_{(k, l)} - v_{(k-1, l+1)}) \\
& - \delta\beta'(z)(1-d_p)k(v_{(k, l)} - v_{(k-1, l+1)})r_{(k, l)}(l-1) \\
& + \delta(v_{(k, l-1)} - v_{(k, l)})r_{(k, l)}(l-1)
\end{aligned}
\tag{1}
$$

where $p_{(i, j)}$ is the equilibrium frequency of class-(i, j) patches, $r_{(k, l)} = r$ (Supplementary Information) is the relatedness between two different parasites on the same class-(k, l) patch, $v_{(k, l)}$ is the reproductive value[17] of a class-(k, l) parasite (the long-term genetic

contribution made by such a parasite) and a prime denotes differentiation. Put verbally, equation (1) shows that the inclusive fitness effects of increased virulence are the cost of killing one's host, the benefits of enhanced transmission, the costs of increased competition for self, the costs of increased competition for relatives and the benefit to relatives due to killing one's host. In Supplementary Information, we show how equation (1) is used to determine the evolutionarily stable level of virulence (z^*).

In clear contrast to recent claims[4–7], analysis of equation (1) shows that the effect of parasite dispersal on virulence can be explained entirely using inclusive fitness theory (Fig. 1). Equation (1) is divided into the direct (personal) fitness consequences of increased virulence (first, second and third lines) and the indirect consequences for relatives (fourth and fifth lines). The first and second lines reflect the assumed compromise between host survival and parasite transmission: the host exploited by the mutant parasite suffers increased mortality (captured by the first term, $-\delta v_{(k, l)}$), whereas the parasite is able to produce new infections—both locally and globally—at a slightly higher rate (captured by

$$
\delta\beta'(z)\left[(1-d_p)kv_{(k-1, l+1)} + d_p\sum\nolimits_{(i, j)} v_{(i, j)}(i+1)p_{(i+1, j-1)}\right]
$$

which is called the 'fecundity change' term). For the special case of a well-mixed parasite population ($d_p = 1$), all but the first line of equation (1) vanishes (Supplementary Information), giving us the standard result that virulence evolves to maximize the basic reproductive number of the parasite[16,22].

The third line of equation (1) describes the direct (personal) fitness consequences, for the mutant, of increased local competition for fewer uninfected hosts. The increased transmissibility of the mutant increases the rate at which uninfected hosts become infected on the mutant's patch: class-(k, l) mutants move to class $(k-1, l+1)$ at a higher rate. Numerical results indicate that $v_{(k, l)} - v_{(k-1, l+1)} > 0$, so this change in the local host population represents an additional direct fitness cost of increased virulence that occurs in structured populations (Fig. 1d). Put simply, if parasite offspring do not disperse, then they decrease the local availability of uninfected hosts and increase the number of parasites competing for them. Consequently, increased parasite dispersal favours higher virulence because it reduces the direct

cost of producing offspring with which the focal individual will have to compete (a reduction in both the difference in reproductive values and in $1 - d_p$ in the third line of equation (1); Fig. 1a,d). This direct cost of parasite offspring production appears to be what has been described as self-shading[12]; thus, the third line of equation (1) can be thought of as a mathematical description of reduced self-shading due to parasite dispersal, but is also analogous to the 'tragedy of the commons'[8].

The fourth and fifth lines of equation (1) describe how the increased virulence exhibited by a mutant also has indirect fitness consequences, through changes to the competitive environment experienced by relatives. The major effect is that the increased transmission that results from the higher virulence of the mutant means that the relatives of the mutant also have increased competition for fewer locally available, uninfected hosts. Line four of equation (1) is simply the third line multiplied by both the number of parasites (other than the mutant actor) on the patch $(l-1)$ and the mean relatedness of those other parasites to the mutant ($r_{(k,l)}$). This indirect cost of increased virulence is reduced by parasite dispersal, through making relatives less likely to interact (Fig. 1b), and by decreasing the extent to which an increased virulence reduces the availability of uninfected hosts to relatives (a reduction in $1 - d_p$, $r_{(k,l)}$ and the difference in reproductive values in the fourth line of equation (1); Fig. 1a, b, d). Consequently, increased parasite dispersal favours higher virulence, because it reduces competition between relatives and, hence, reduces the indirect cost of higher virulence. This is analogous to self-shading but applies to the relatives of the mutant actor, so it could be thought of as kin shading. Kin shading is a between-hosts equivalent to the previous result that a lower relatedness (higher strain diversity) within the same host favours higher parasite virulence because it selects for faster growth rates, to obtain a higher proportion of the host resources[16].

The fifth line of equation (1) shows that the increased host mortality due to increased virulence affects the competitive environment experienced by the relatives of the mutant. Increased host mortality benefits relatives because it leads to a reduction in the number of locally competing parasites (a parasite dies along with its host) and because it clears a space that can later be filled by newborn (uninfected) hosts. Increased host mortality is also potentially costly to relatives, because it reduces the number of local hosts (a source of newborn, uninfected hosts). In many cases, numerical results indicate that $v_{(k,l-1)} - v_{(k,l)} > 0$, meaning that the increased host mortality that results from increased mutant virulence provides a net benefit to the mutant's relatives (upper lines in Fig. 1d). In these same cases, reduced parasite dispersal leads to a decrease in the competition experienced by relatives and, hence, an indirect benefit to higher virulence. In other cases, $v_{(k,l)} > v_{(k,l-1)}$ and the term $\delta(v_{(k,l-1)} - v_{(k,l)})r_{(k,l)}(l-1)$ of equation (1) counts as a cost in increased virulence. These latter cases are characterized by low host dispersal rates, so the cost of increased virulence here stems from the depletion of the main source of new, uninfected hosts. We must emphasize that even when $\delta(v_{(k,l-1)} - v_{(k,l)})r_{(k,l)}(l-1)$ counts as an inclusive fitness benefit, its size at equilibrium appears to be insufficient to raise the evolutionarily stable virulence level above that found in well-mixed populations.

More generally, as well as clarifying why the parasite dispersal rate should influence virulence, our model also shows how and why the parasite dispersal rate will interact with other parameters such as the maximum transmissibility of the parasite (β_{max}), the reproductive rate of the host and the host dispersal rate (Fig. 2). Increased host dispersal would favour increased virulence, through decreasing the extent to which increased virulence leads to self-shading and kin competition, as well as through any influence on parasite dispersal by moving parasites within hosts.

The reason why the parasite dispersal rate should influence virulence has proved controversial. Previous studies have offered four different explanations: virulent strains being surrounded by other infected individuals (self-shading)[12]; over-exploitation of the local availability of hosts (tragedy of the commons)[8]; competition between related strains ('kin selection')[15]; and the over-exploitation of local hosts and, hence, the extinction of parasite groups (the original Wynne-Edwards theory of 'group selection')[6,11]. It has also been suggested that the relationship between parasite dispersal and virulence is beyond the scope of existing evolutionary theory[25].

Our results show that an increase in parasite dispersal rate leads to selection for increased growth and, hence, to higher virulence for three reasons (Fig. 1). Increased dispersal provides a direct benefit to greater virulence, because it (1) increases the relative value of producing offspring (dispersers are worth more than non-dispersers) and (2) reduces the extent to which producing offspring will lead to the focal individual experiencing an increase in competition for available hosts (self-shading[12]). Increased dispersal provides an indirect benefit to greater virulence, because it (3) reduces the extent to which producing offspring will lead to relatives experiencing an increase in competition for available hosts (kin shading). The previous verbal explanations can be linked to these causal forces, in that self-shading[12] is our reason 1, the tragedy of the commons[8] involves our reasons 2 and 3, competition between relatives[15] is our reason 3 and the extinction of parasite groups[6,11] is linked to reasons 2 and 3; if an individual causes harm to their patch, then this cost is paid by both the focal individual and their relatives on the patch (that is, the group-selection components can always be partitioned into offspring and non-offspring components). There is also a fourth effect that works in the opposite direction, favouring a lower virulence with increased parasite dispersal. This effect is a consequence of the indirect benefit of reduced competition due to the number of relatives dying being greater at lower dispersal rates, but this is outweighed by the other three factors.

To conclude, we have shown that selection on rare mutant virulence phenotypes in structured populations of parasites can be explained by inclusive fitness theory. This is the latest of numerous
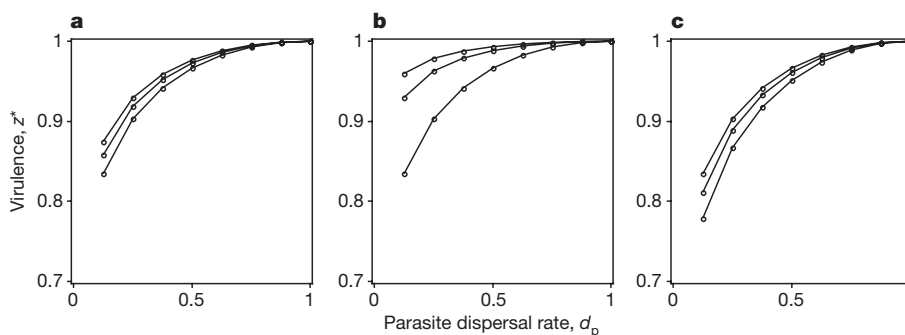

Figure 2 | Host and parasite life histories affect the relationship between stable virulence level and parasite dispersal rate. Relationship between z^* and d_p as host life-history parameters vary (**a**, **b**) and as maximum disease transmissibility, a parasite life-history trait, varies (**c**). **a**, From top to bottom, $d_h = 0.9$, 0.6 and 0.3 ($b = 3$). **b**, From top to bottom, $b = 9$, 6 and 3 ($d_h = 0.3$). Remaining parameters in **a** and **b** were $\mu = 1$, $N = 5$ and $\beta(x) = 5x/(1 + x)$. **c**, From top to bottom, $\beta_{max} = 5$, 7.5 and 20; remaining parameters were $\mu = 1$, $d_h = 0.3$, $b = 3$, $N = 5$ and $\beta(x) = \beta_{max}x/(1 + x)$. Additional, qualitatively similar results are presented in Supplementary Information.

985

examples that have accumulated over the past 30 years, in which it has been claimed that group selection and not kin selection is acting in a particular situation, only for explicit analyses to show otherwise[26,27]. Future confusion could be avoided if such claims were backed by formal analyses that actually examine the underlying selective forces, rather than just verbal arguments[26]. More generally, our results emphasize the difference between levels of adaptation and levels of selection[28]. The multilevel (group) selection and kin selection (inclusive fitness) approaches to social evolution have long been known to be mathematically equivalent and, if the analyses are performed correctly, do not lead to conflicting predictions[29,30]. Thus, irrespective of the relative strengths of within-group versus between-group selection, individuals are predicted to maximize their inclusive fitness. In contrast, groups are only predicted to evolve traits that function to maximize their fitness in extreme situations where there is no conflict of interest between the members of the group[28]. Put another way, the presence of group selection does not invalidate the idea that the individual is an adaptive unit, and it does not validate the idea that the group is an adaptive unit[28].

1. Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–52 (1964).
2. Grafen, A. Optimization of inclusive fitness. *J. Theor. Biol.* **238**, 541–563 (2006).
3. Grafen, A. The formal Darwinism project: a mid-term report. *J. Evol. Biol.* **20**, 1243–1254 (2007).
4. Sober, E. & Wilson, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, 1998).
5. Wilson, D. S. & Wilson, E. O. Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.* **82**, 327–348 (2007).
6. Wilson, D. S. Social semantics: towards a genuine pluralism in the study of social behaviour. *J. Evol. Biol.* **21**, 368–373 (2008).
7. Hölldobler, B. & Wilson, E. O. *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* (Norton, 2008).
8. Kerr, B., Neuhauser, C., Bohannan, J. M. & Dean, A. M. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature* **442**, 75–78 (2006).
9. Boots, M. & Mealor, M. Local interactions select for lower pathogen infectivity. *Science* **315**, 1284–1286 (2007).
10. Wynne-Edwards, V. C. *Animal Dispersion in Relation to Social Behaviour* (Oliver & Boyd, 1962).
11. Haraguchi, Y. & Sasaki, A. Host-parasite arms race in mutation modifications: indefinite escalation despite a heavy load. *J. Theor. Biol.* **183**, 121–137 (1996).
12. Boots, M. & Sasaki, A. 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B* **266**, 1933–1938 (1999).
13. O'Keefe, K. J. & Antonovics, J. Playing by different rules: the evolution of virulence in sterilizing pathogens. *Am. Nat.* **159**, 597–605 (2002).
14. Boots, M., Hudson, P. J. & Sasaki, A. Large shifts in pathogen virulence relate to host population structure. *Science* **303**, 842–844 (2004).
15. Lion, S. & van Baalen, M. Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* **11**, 277–295 (2008).
16. Frank, S. A. Models of parasite virulence. *Q. Rev. Biol.* **71**, 37–78 (1996).
17. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, 1930).
18. Hamilton, W. D. Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220 (1970).
19. Stearns, S. C. *Evolution of Life Histories* (Oxford Univ. Press, 1992).
20. Krebs, J. R. & Davies, N. B. *An Introduction to Behavioural Ecology* 3rd edn (Blackwell Scientific, 1993).
21. Rousset, F. & Ronce, O. Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* **65**, 127–141 (2004).
22. Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85**, 411–426 (1982).
23. Day, T. & Gandon, S. Applying population-genetic models in theoretical evolutionary epidemiology. *Ecol. Lett.* **10**, 876–888 (2007).
24. Rousset, F. *Genetic Structure and Selection in Subdivided Populations* (Princeton Univ. Press, 2004).
25. Goodnight, C. *et al.* Evolution in spatial predator–prey models and the "prudent predator": the inadequacy of steady-state organism fitness and the concept of individual and group selection. *Complexity* **13**, 23–44 (2008).
26. Lehmann, L. & Keller, L. The evolution of cooperation and altruism – a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376 (2006).
27. West, S. A., Griffin, A. S. & Gardner, A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432 (2007).
28. Gardner, A. & Grafen, A. Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671 (2009).
29. Frank, S. A. Hierarchical selection theory and sex ratios. I. General solutions for structured populations. *Theor. Popul. Biol.* **29**, 312–342 (1986).
30. Queller, D. C. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* **139**, 540–558 (1992).

LETTERS

Copy number variation at 1q21.1 associated with neuroblastoma

Sharon J. Diskin[1,2], Cuiping Hou[1,3], Joseph T. Glessner[3], Edward F. Attiyeh[1,4], Marci Laudenslager[1], Kristopher Bosse[1], Kristina Cole[1], Yaël P. Mossé[1,4], Andrew Wood[1], Jill E. Lynch[1], Katlyn Pecor[1], Maura Diamond[1], Cynthia Winter[1], Kai Wang[3], Cecilia Kim[3], Elizabeth A. Geiger[6], Patrick W. McGrady[8], Alexandra I. F. Blakemore[9], Wendy B. London[8], Tamim H. Shaikh[4,6], Jonathan Bradfield[3], Struan F. A. Grant[3,4,6], Hongzhe Li[7], Marcella Devoto[4,6,7,10], Eric R. Rappaport[1,4], Hakon Hakonarson[3,4,6] & John M. Maris[1,4,5]

Common copy number variations (CNVs) represent a significant source of genetic diversity, yet their influence on phenotypic variability, including disease susceptibility, remains poorly understood. To address this problem in human cancer, we performed a genome-wide association study of CNVs in the childhood cancer neuroblastoma, a disease in which single nucleotide polymorphism variations are known to influence susceptibility[1,2]. We first genotyped 846 Caucasian neuroblastoma patients and 803 healthy Caucasian controls at ~550,000 single nucleotide polymorphisms, and performed a CNV-based test for association. We then replicated significant observations in two independent sample sets comprised of a total of 595 cases and 3,357 controls. Here we describe the identification of a common CNV at chromosome 1q21.1 associated with neuroblastoma in the discovery set, which was confirmed in both replication sets. This CNV was validated by quantitative polymerase chain reaction, fluorescent *in situ* hybridization and analysis of matched tumour specimens, and was shown to be heritable in an independent set of 713 cancer-free parent–offspring trios. We identified a previously unknown transcript within the CNV that showed high sequence similarity to several neuroblastoma breakpoint family (NBPF) genes[3,4] and represents a new member of this gene family (*NBPF23*). This transcript was preferentially expressed in fetal brain and fetal sympathetic nervous tissues, and the expression level was strictly correlated with CNV state in neuroblastoma cells. These data demonstrate that inherited copy number variation at 1q21.1 is associated with neuroblastoma and implicate a previously unknown neuroblastoma breakpoint family gene in early tumorigenesis of this childhood cancer.

Neuroblastoma is a paediatric cancer of the developing sympathetic nervous system that most commonly affects young children, and is often lethal[5]. Most neuroblastomas arise sporadically, with less than 1% of cases inherited in an autosomal dominant fashion[5]. We recently identified the anaplastic lymphoma kinase (*ALK*) gene as the major hereditary neuroblastoma predisposition gene[1]. For the majority of neuroblastomas that arise without a family history of the disease, we hypothesize that multiple common DNA variations cooperate to increase the risk for neuroblastic malignant transformation. By performing a genome-wide association study (GWAS) of single nucleotide polymorphism (SNP) genotypes, we recently identified

common SNP alleles within the putative *FLJ22536* and *FLJ44180* genes at 6p22 and within the *BARD1* gene at 2q35 associated with malignant neuroblastoma, providing the first evidence that childhood cancers also arise owing to complex interactions of polymorphic variants[1,2]. Here we investigate constitutional DNA CNVs as another source of genetic diversity that may contribute to the development of this disease. CNVs have been shown to significantly influence messenger RNA expression levels[6] and recent studies have described associations of CNVs with systemic autoimmunity[7,8], autism[9], schizophrenia[10–12] and psoriasis[13]. In addition, an increased number of CNVs in Li–Fraumeni families harbouring *TP53* mutations has been reported, but the associations of individual CNVs with cancer susceptibility have not been explored[14].

To identify CNVs that are associated with neuroblastoma, we first genotyped a discovery set of 1,032 Caucasian neuroblastoma patients and 2,043 disease-free Caucasian control subjects, as previously described[1,15]. We next applied stringent quality control criteria necessary for accurate CNV detection and reliable association assessment (see Supplementary Methods). The final discovery set consisted of 846 Caucasian cases and 803 Caucasian controls (Supplementary Tables 1 and 2). These subjects showed tight clustering in a multidimensional scaling analysis of SNPs not in linkage disequilibrium (Supplementary Fig. 1), demonstrating that population substructure was not likely to have a significant impact on association testing.

By comparing single marker binary copy number states at 531,689 SNPs mapped to autosomes, we observed a total of 131 SNPs showing significant association with neuroblastoma, defined as a two-sided Fisher's exact test P value below a genome-wide threshold of $P = 1.0 \times 10^{-7}$ (Supplementary Table 3). Associations with deletion polymorphisms were seen at chromosomes 1, 7 and 14 (Fig. 1a); no duplication polymorphisms reached genome-wide significance (Fig. 1b). Review of significant SNPs revealed four distinct regions of deletion (Table 1). We next sought to validate significant association signals in two independent replication sets, the first consisting of 363 Caucasian cases and 1,139 Caucasian controls, and the second of 232 Caucasian cases and 2,218 Caucasian controls (Supplementary Tables 1 and 2). All deletion associations showed robust replication in both independent case series (Table 2 and Supplementary Tables 4 and 5).

We observed a seven-SNP deletion at 1q21.1 that occurred in 15.6% of cases but only 9.1% of controls overall ($P_{combined} = 2.97 \times 10^{-17}$;

[1]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, [2]Genomics and Computational Biology, University of Pennsylvania School of Medicine, [3]The Center for Applied Genomics, Children's Hospital of Philadelphia, [4]Department of Pediatrics, University of Pennsylvania School of Medicine, [5]Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, [6]Division of Genetics, The Children's Hospital of Philadelphia, [7]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, 19104, USA. [8]Department of Statistics, University of Florida and Children's Oncology Group, Gainesville, Florida, 32611, USA. [9]Genomic Medicine, Imperial College London, London, SW7 2AZ, UK. [10]University of Rome "La Sapienza", Department of Experimental Medicine, Rome, 00185, Italy.
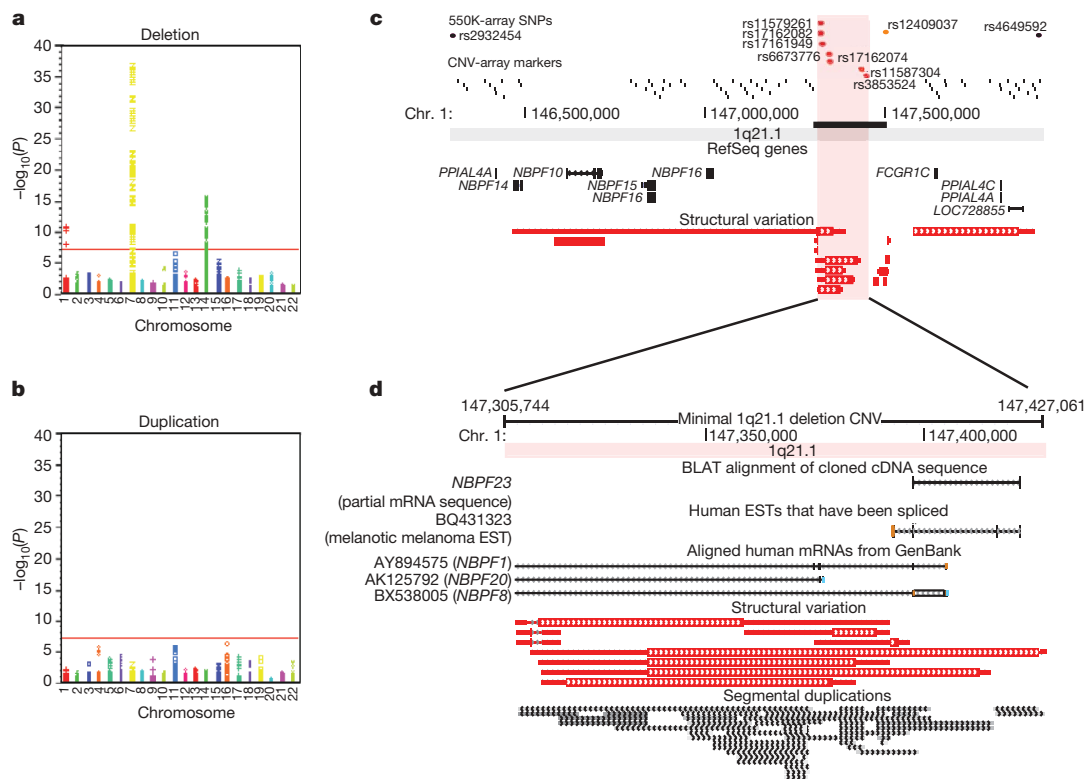
Figure 1 | Discovery of 1q21.1 CNV associated with neuroblastoma.
a, Summary of discovery results for deletions. Level of significance ($\log_{10}$ transformed P values) for each SNP at each chromosome is plotted. Red line, genome-wide significance threshold. **b**, Summary of discovery results for duplications. **c**, Genomic map and refinement of 1q21.1 deletion boundaries. SNPs from the 550K array and markers from the CNV-12 array are plotted as dots or tick marks; those reaching genome-wide significance are coloured red. Image borders represent maximal deletion boundaries based on 550K array data, and include a cluster of NBPF genes and several other RefSeq transcripts. Refined maximal boundaries using CNV-12 data (black bar) are consistent with published structural variations. **d**, Genomic map of 121-kb minimal 1q21.1 deletion covered by significant SNPs. The region contains a high level of segmental duplication that extends to flanking regions (not shown). The BQ431323 melanoma EST maps uniquely within this region, whereas several GenBank transcripts align to the region with >95% identity but map elsewhere on the genome. BLAT alignment of PCR amplified and cloned sequence from fetal brain in this study is shown (NBPF23).

odds ratio (OR), 2.49; 95% confidence interval (CI), 2.02–3.05). This association remained significant after additional adjustment for potential population substructure captured by SNPs not in linkage disequilibrium ($P_{\text{discovery}} < 0.0001$), and was driven by a difference in hemizygous deletion frequency ($P_{\text{combined}} = 1.83 \times 10^{-19}$). The observed frequency of homozygous deletion was identical in cases and controls (1.3% overall). The maximal deletion defined by neighbouring two-copy SNPs spanned 1.6 Mb and contained a cluster of NBPF genes (Fig. 1c). To reduce the maximal deletion boundaries, we first genotyped 48 representative samples on the Illumina CNV-12 array (36 cases and 12 controls equally divided between those with and without the deletion CNV). These data refined the maximal size of the deletion to approximately 300 kb, consistent with published reports of CNVs at this location (Fig. 1c)[16–18]. Finally, use of the Illumina HumanHap610 SNP platform reduced the maximal deletion

size to only 143 kb, from 147,292,384–147,435,422 bp (Supplementary Fig. 2). The minimal deletion based on significant SNPs in the association study spanned 121 kb from 147,305,744–147,427,061 bp and did not contain any known genes (Fig. 1d).

We next sought to confirm that this CNV is indeed a heritable genetic variation. First we genotyped an independent set of 713 trios from variable phenotypes on the same 550K SNP array and generated CNV calls using a family-based approach[16]. Deletion at 1q21.1 was observed in 125 offspring and confirmed by parental analysis in 123 trios, estimating the inheritance rate to be 98.4%. Next we genotyped paired tumour DNA in 226 cases (Supplementary Table 2) using the same 550K SNP platform, and confirmed existence of the CNV (deletion and duplication) in every tumour sample studied (Supplementary Fig. 3a). We did not observe progression from hemizygous deletion in constitutional DNA to homozygous deletion

Table 1 | Significant copy number variable regions from the discovery phase

Chr.	Start SNP / End SNP	Start position / End position	No. of SNPs	Size (bp)	Gene(s)	Case % loss (range)	Control % loss (range)	Case % gain (range)	Control % gain (range)	P-value	OR (95% CI)
1	rs11579261 / rs3853524	147,305,744 / 147,427,061	7	121,317	NBPF23	15.1 (12.2–15.1)	5.2 (3.6–5.2)	3.2 (1.8–3.2)	5.1 (1.6–5.1)	2.38×10^{-11}	3.23 (2.25–4.05)
7	rs7782269 / rs733905	38,285,115 / 38,346,971	47	61,856	TRG@	24.5 (8.6–36.4)	7.3 (0.5–16.2)	0	0.1 (0.1–0.1)	2.20×10^{-37}	9.96 (6.40–15.48)
7	rs2213212 / rs2367486	142,086,318 / 142,192,134	11	105,816	TRB@	4.6 (3.9–5.2)	0.2 (0.1–0.3)	0.1 (0.0–0.1)	0	5.80×10^{-11}	21.97 (5.85–82.42)
14	rs979027 / rs2128997	21,558,349 / 22,030,942	66*	472,593	TRA@, TRD@	29.2 (5.2–45.1)	14.1 (0.6–30.3)	0.0 (0.0–0.1)	0.0 (0.0–0.1)	2.99×10^{-16}	2.71 (2.12–3.46)

Genomic coordinates are based on UCSC Build 36.1 of the human genome. The P value is based on two-tailed Fisher's exact test comparing deletion frequency in cases versus controls (minimum P value listed; see Supplementary Information for full range). Median percentage deletion and duplication are listed, with the range in parentheses. ORs listed are from the most significant SNP in the deleted region: 1q21.1 (NBPF23), rs17162082; TRG@, rs718880; TRB@, rs6959895; TRA@/TRD@, rs741711. n = 846 for cases and n = 803 for controls.
* SNPs not consecutive.

Table 2 | Replication of significant copy number variable regions

Chr.	Start SNP End SNP	Start position End position	No. of SNPs	Size (bp)	Gene	Case % loss (range)	Control % loss (range)	Case % gain (range)	Control % gain (range)	Replication P-value	Combined P-value	Combined OR (95% CI)
Replication set 1												
1	rs11579261 rs3853524	147,305,744 147,427,061	7	121,317	NBPF23	14.3 (10.7–14.3)	8.7 (6.4–8.7)	1.9 (1.7–1.9)	3.1 (1.9–3.1)	2.58×10^{-3}	1.70×10^{-11}	2.23 (1.77–2.82)
7	rs7782269 rs733905	38,285,115 38,346,971	47	61,856	TRG@	14.9 (5.8–30.3)	3.8 (0.2–9.3)	0	0	8.61×10^{-23}	1.17×10^{-63}	10.74 (7.73–14.93)
7	rs2213212 rs2367486	142,086,318 142,192,134	11	105,816	TRB@	3.9 (1.9–4.4)	0	0	0	1.05×10^{-10}	8.59×10^{-23}	46.23 (12.40–172.34)
14	rs979027 rs2128997	21,558,349 22,030,942	66*	472,593	TRA@, TRD@	24.7 (2.2–30.9)	11.6 (0.3–19.9)	0	0	1.47×10^{-11}	3.54×10^{-32}	2.82 (2.35–3.38)
Replication set 2												
1	rs11579261 rs3853524	147,305,744 147,427,061	7	121,317	NBPF23	19.0 (14.2–19.0)	10.6 (5.4–10.7)	3.5 (2.6–3.5)	4.1 (3.4–4.1)	2.59×10^{-6}	2.97×10^{-17}	2.49 (2.02–3.05)
7	rs7782269 rs733905	38,285,115 38,346,971	47	61,856	TRG@	19.0 (7.7–31.0)	3.1 (0.4–7.3)	0	0	2.22×10^{-23}	1.64×10^{-103}	12.62 (9.74–16.36)
7	rs2213212 rs2367486	142,086,318 142,192,134	11	105,816	TRB@	3.0 (3.0–4.3)	0.1 (0.1–0.2)	0	0.0 (0.0–0.1)	1.07×10^{-8}	2.40×10^{-35}	41.79 (16.83–103.80)
14	rs979027 rs2128997	21,558,349 22,030,942	66*	472,593	TRA@, TRD@	22.2 (4.3–34.1)	7.2 (0.4–14.5)	0	0.0 (0.0–0.1)	6.25×10^{-15}	3.19×10^{-63}	3.75 (3.22–4.36)

See Table 1 for details. Replication set 1: $n = 363$ for cases and $n = 1,139$ for controls. Replication set 2: $n = 232$ for cases and $n = 2,218$ for controls.
* SNPs not consecutive.

in the matched tumour DNA for any case in this study, nor did we observe any expansion of CNV boundaries in the tumour compared to matched constitutional DNA.

To investigate whether 1q21.1 deletions are associated with specific neuroblastoma phenotypes, we tested each for association with clinical covariates using the combined set of 1,441 cases (Supplementary Table 6). Although deletions at 1q21.1 were observed more frequently in patients with aggressive disease, this trend did not reach statistical significance in this study. An additive effect on the odds ratio was observed for those harbouring both the 1q21.1 deletion and the 6p22 risk alleles (Supplementary Fig. 4), but no significant interaction effect was detected to indicate epistasis (Supplementary Table 7).

In addition to the 1q21.1 CNV, we observed highly significant associations of deletion within all four T-cell receptor (TCR) loci clustered on chromosomes 7 and 14. Contrary to the 1q21.1 CNV, these deletions were not observed in paired tumour DNA samples (Supplementary Fig. 3b). Heterogeneity was observed in the areas of apparent deletion, consistent with some cells harbouring homozygous deletion and others exhibiting two-copy heterozygosity. Deletions within different TCRs tended to co-occur in patients ($P < 0.0001$, Supplementary Table 7), and were significantly more common in blood- than bone-marrow-derived DNA samples ($P_{TRG@} = 9.2 \times 10^{-31}$, $P_{TRB@} = 7.3 \times 10^{-6}$, $P_{TRA@/TRD@} = 8.0 \times 10^{-50}$). Taken together, these findings indicate that we are detecting an oligoclonal expansion of T-cell lymphocytes in a subset of neuroblastoma patients, and this signal is diluted within the bone marrow compartment. Interestingly, TCR deletions showed a marked over-representation in the less aggressive subset of neuroblastoma (Supplementary Table 6 and Supplementary Fig. 5). It is possible that these events herald an immunologic response to neuroblastoma; however, this hypothesis requires further investigation.

To validate the 1q21.1 CNV, we first performed quantitative polymerase chain reaction (qPCR) on 46 neuroblastoma cases (10 zero-copy, 16 one-copy, 12 two-copy and 8 duplications as predicted by SNP analysis). We observed 100% concordant results when comparing copy number estimated by qPCR with copy number based on SNP genotyping (Fig. 2a). To confirm that the detected CNV is not an artefact caused by segmental duplication of the 1p36 region, and that it indeed maps to 1q21.1, we validated the existence of the deletion in a sample harbouring a single-copy loss using fluorescent *in situ* hybridization (Fig. 2b).

Although no known genes mapped to the refined 1q21.1 CNV, we identified a spliced EST (BQ431323) from a melanoma library that mapped within the CNV with 100% identity across the entire sequence (Fig. 1d). Using primers designed against exon 1 and exon 3 of BQ431323, we PCR-amplified complementary DNA from fetal brain and a neuroblastoma cell line. These PCR products were cloned and sequenced (GenBank accession number GQ131873); the resulting sequence mapped uniquely within the CNV with 100% identity across the entire sequence and showed splicing out of the predicted second exon of BQ431323 (Supplementary Fig. 6). The top scoring hit from a Blastn[19] search of this sequence against available human RefSeq transcripts was NBPF3 ($E = 5.0 \times 10^{-89}$), followed by NBPF1 ($E = 1.0 \times 10^{-75}$) and NBPF15 ($E = 2.0 \times 10^{-24}$). These data provide strong evidence for a new NBPF transcript, named NBPF23 here, mapping within the 1q21.1 CNV associated with neuroblastoma.

We investigated the expression of NBPF23 in both neuroblastoma cells and normal human fetal and adult tissues using real-time quantitative reverse transcriptase PCR. Analysis of 18 neuroblastomas (tumours and cell lines) of known copy number at the 1q21.1 CNV showed a clear correlation between CNV state and transcript expression (Fig. 2c). Notably, two-copy samples clustered into two distinct expression classes ($P = 0.007$), the first with low expression and the second with high expression, and these probably represent different 1q21.1 CNV genotypes. There are two possible CNV genotypes for two-copy samples, which we refer to as 2:0 ('cis') and 1:1 ('trans') based on the number of copies present on each chromosome in a diploid genome (Supplementary Fig. 7a). Two neuroblastoma samples in the low-expression group can be demonstrated to be from the 2:0 constitutional CNV genotype because they have somatically acquired gain of chromosome 1q (three copies) yet are two copies with heterozygous SNPs at the 1q21.1 CNV (see Supplementary Fig. 7b, c for details). These findings are consistent with the hypothesis that two copies in 'cis' (same chromosome) behave differently to two copies in 'trans' (different chromosomes). Therefore, we propose a model whereby NBPF23 expression is decreased when copies are in the cis configuration as opposed to the trans configuration. Importantly, however, even when the two-copy samples are not clustered in this manner, a statistically significant difference in transcript levels is observed between the two- and three-copy samples ($P = 0.001$). Finally, we analysed expression of NBPF23 in a panel of 28 normal fetal and adult tissues (Fig. 2d). We observed the highest transcript levels in fetal brain and fetal sympathetic ganglia from early gestation (13–22 weeks), consistent with NBPF23 being expressed in early sympathicoadrenal neurodevelopment.

NBPF genes were identified after the founding member, NBPF1, was determined to be disrupted by means of a constitutional chromosomal translocation in a neuroblastoma patient[3,4]. Subsequent scans of the genome identified three clusters of NBPF genes on chromosome 1 within areas of segmental duplication. The encoded proteins are recently evolved and primate-specific, share significant homology and contain highly conserved domains of unknown function
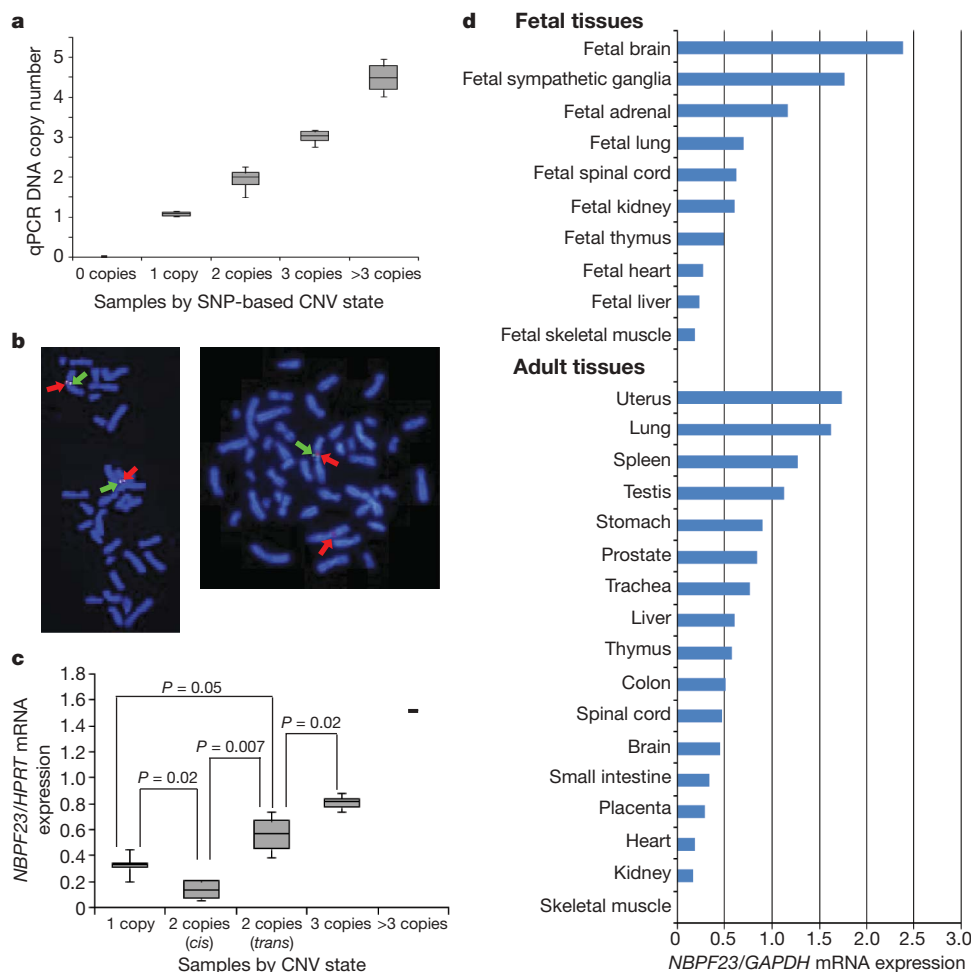
Figure 2 | Validation and biological relevance of 1q21.1 CNV.
a, Quantitative PCR validation of 1q21.1 CNV states shows 100% agreement with SNP genotyping data. Box and whisker plot is shown for 0 copies ($n = 10$), 1 copy ($n = 16$), 2 copies ($n = 12$), 3 copies ($n = 6$) and >3 copies ($n = 2$). Lower and upper whiskers (error bars) indicate 10th and 90th percentiles respectively. **b**, Fluorescent *in situ* hybridization (FISH) validation of 1q21.1 deletion. Shown are metaphase spreads from lymphoblastoid cell lines established from an individual with the normal two copies (left) and an individual with hemizygous deletion (right). Red arrows: BAC RP4-790G17, which maps just outside of the deletion CNV from 148,424,830–148,557,655 bp (two copies in both individuals). Green arrows: fosmid W12-1967b11, which maps within the deleted region from 147,382,742–147,420,968 bp (one copy in individual with hemizygous deletion). **c**, Box and whisker plot showing correlation between CNV state and *NBPF23* transcript levels demonstrated using real-time quantitative reverse transcriptase PCR in 18 neuroblastoma samples: 1 copy ($n = 5$), 2 copies (low, $n = 4$), 2 copies (high, $n = 3$), 3 copies ($n = 5$) and >3 copies ($n = 1$). Lower and upper whiskers (error bars) indicate 10th and 90th percentiles respectively. Two-copy samples clustered into two distinct groups consistent with expression difference based on whether copies are in *cis* (2:0 CNV genotype) or *trans* (1:1 CNV genotype); see text for details. **d**, Expression of *NBPF23* in normal fetal and adult tissues using real-time quantitative reverse transcriptase PCR showed highest levels in fetal brain (13–22 weeks) and fetal sympathetic ganglia samples (18–22 weeks), which were collected at the Children's Hospital of Philadelphia.

(DUF1220) that are thought to be neuronal-specific[3,20]. Constitutional deletions disrupting NBPF genes have been implicated in schizophrenia[10–12] and autism[10], and rare recurrent structural variants just upstream of the CNV identified in this study have also been reported in a variety of phenotypes including mental retardation, autism and congenital anomalies[21]. Although the specific functions of NBPF genes are not known, recently evolved genes may be involved in cancer predisposition owing to a lack of selective pressure[12]. Expression of *NBPF1* was recently shown to suppress anchorage-independent growth[4]. Conversely, overexpression of NBPF transcripts has been reported in several cancers including sarcomas[22] and non-small-cell lung cancer[23]. A challenge for the field is to design experiments that clearly distinguish these highly homologous transcripts and define specific disease-causal mechanisms.

Neuroblastoma develops from the malignant transformation of partially committed sympathicoadrenal neuroblasts during fetal or early childhood development. We previously identified the *ALK* gene as the major familial neuroblastoma predisposition gene[1] and showed that common SNP variations at 6p22 and 2q35 are associated with sporadic neuroblastma[1,2]. In our current study, we show that a common CNV at 1q21.1 likewise contributes to neuroblastoma susceptibility, and that this CNV leads to altered expression of *NBPF23*, a previously unknown NBPF transcript. These data provide evidence for a specific CNV predisposing to human cancer, and ongoing efforts will define remaining susceptibility variants in the human genome associated with neuroblastoma.

METHODS SUMMARY

A multi-phase GWAS of CNVs in neuroblastoma was performed. A discovery set of 846 Caucasian neuroblastoma patients and 803 Caucasian healthy controls were genotyped using a 550K SNP array and CNV profiles were generated[16,24], providing a copy number state for each SNP on the array. Association with neuroblastoma was assessed at 531,689 individual SNPs on autosomes using Fisher's exact test on the binary comparison of 'deletion versus no deletion' or 'duplication versus no duplication'. A genome-wide significance threshold was set at $P = 1.0 \times 10^{-7}$ and replication of significant findings was assessed using two independent replication sets, the first consisting of 363 Caucasian neuroblastoma cases and 1,139 Caucasian controls and the second of 232 Caucasian neuroblastoma cases and 2,218 Caucasian controls. Matched tumour DNA from

226 cases was genotyped to evaluate whether observed deletions were constitutional or somatically acquired rearrangements in the blood. An independent set of 713 trios with no reported incidence or history of cancer was genotyped and used to establish heritability of significant CNVs. Inherited CNVs were validated by quantitative real-time PCR and fluorescent *in situ* hybridization. Association with clinical and biological co-variates was assessed on the combined set of 1,441 neuroblastoma cases with available annotation. PCR amplification of cDNA from fetal brain and neuroblastoma cell lines was performed to identify a novel 1q21.1 transcript; products from these PCR reactions were sequenced and aligned to the human genome using BLAT[25]. Similarity to known *NBPF* transcripts was detected using Blastn[19]. Expression of the novel *NBPF* transcript was assessed using real-time quantitative reverse transcriptase PCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Maris, J. M. *et al.* Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N. Engl. J. Med.* **358**, 2585–2593 (2008).
2. Capasso, M. *et al.* Common variations in *BARD1* influence susceptibility to high-risk neuroblastoma. *Nature Genet.* doi:10.1038/ng.374 (3 May 2009).
3. Vandepoele, K., Van Roy, N., Staes, K., Speleman, F. & Van Roy, F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol. Biol. Evol.* **22**, 2265–2274 (2005).
4. Vandepoele, K. *et al.* A constitutional translocation t(1;17)(p36.2;q11.2) in a neuroblastoma patient disrupts the human *NBPF1* and *ACCN1* genes. *PLoS ONE* **3**, e2207 (2008).
5. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
6. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
7. Aitman, T. J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
8. Fanciulli, M. *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature Genet.* **39**, 721–723 (2007).
9. Sebat, J. *et al.* Stong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
10. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
11. Stone, J. L. *et al.* Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
12. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
13. Hollox, E. J. *et al.* Psoriasis is associated with increased b-defensin genomic copy number. *Nature Genet.* **40**, 23–25 (2008).
14. Shlien, A. *et al.* Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc. Natl Acad. Sci. USA* **105**, 11264–11269 (2008).
15. Steemers, F. J. *et al.* Whole-genome genotyping with the single-base extension assay. *Nature Methods* **3**, 31–33 (2006).
16. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
17. Conrad, D. F., Andrews, T. D., Carter, N., Hurles, M. & Pritchard, J. K. A high resolution survey of deletion polymorphisms in the human genome. *Nature Genet.* **38**, 75–81 (2006).
18. Pinto, D., Marshall, C., Feuk, L. & Scherer, S. W. Copy-number variation in control population cohorts. *Hum. Mol. Genet.* **2**, R168–R173 (2007).
19. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
20. Popesco, M. C. *et al.* Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307 (2006).
21. Mefford, H. C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
22. Meza-Zepeda, L. A. *et al.* Positional cloning identifies a novel cyclophilin as a candidate amplified oncogene in 1q21. *Oncogene* **21**, 2261–2269 (2002).
23. Petroziello, J. *et al.* Suppression subtractive hybridization and expression profiling identifies a unique set of genes overexpressed in non-small-cell lung cancer. *Oncogene* **23**, 7734–7745 (2004).
24. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36**, e126 (2008).
25. Kent, W. J. BLAT — the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

METHODS

Sample selection. Cases were defined as a child diagnosed with neuroblastoma or ganglioneuroblastoma and registered through the Children's Oncology Group (COG). All specimens were obtained at the time of diagnosis and most were annotated with clinical and genomic information that included: age at diagnosis, site of origin, disease stage by the International Neuroblastoma Staging System[26], INPC International Neuroblastoma Pathology Classification[27], *MYCN* oncogene copy number[28], DNA index (ploidy)[29], registration on clinical trial(s), event-free and overall survival, second malignancies and any associated conditions (for example, congenital abnormalities). Additional eligibility criteria and quality control measures, including exclusion of samples with evidence of circulating tumour DNA, are detailed in Supplementary Methods.

Control subjects were recruited from the Philadelphia region through the CHOP Health Care Network, including four primary care clinics and several group practices and outpatient practices that included well child visits. Eligibility criteria for control subjects were: self-reported as Caucasian, availability of 1.5 μg of high quality DNA from peripheral blood mononuclear cells and no serious underlying medical disorder, including cancer.

Genome-wide SNP genotyping. Genotyping for both discovery and replication phases was performed using the Illumina Infinium II HumanHap550 BeadChip according to methods detailed elsewhere[15,30] and summarized in Supplementary Methods.

CNV detection. Log R ratio signal intensities were adjusted for 'wave-like' artefacts using a regression-based approach described previously[24] and PennCNV[16] was used to call CNVs in each sample. The copy number at each SNP was then inferred based on projection of CNVs onto the genomic sequence.

Statistical tests. The single-marker statistical analysis for the genome-wide data was carried out using Fisher's exact test on binary copy number differences (deletion versus no deletion or duplication versus no duplication) between cases and controls. Two-sided P values were reported and the OR and corresponding 95% CIs were calculated for genome-wide significant copy number association results. A threshold of 1.0×10^{-7} was set for genome-wide significance based on the fact that approximately 500,000 SNPs were tested ($0.05/500,000 = 1.0 \times 10^{-7}$).

CNVs reaching genome-wide significance were tested for association with clinical and biological co-variates using Fisher's exact test. All comparisons were made based on binary comparison of 'deletion versus no deletion' or 'duplication versus no duplication', and two-sided P values were reported. For deletions within the TCRs, only cases from peripheral blood (not bone marrow) were considered in the clinical correlative analyses because we observed a significant association of TCR deletions with blood versus bone marrow ($P_{TRG@} = 9.2 \times 10^{-31}$, $P_{TRB@} = 7.3 \times 10^{-6}$, $P_{TRA@/TRD@} = 8.0 \times 10^{-50}$).

CNV heritability. We used Illumina HumanHap550 SNP genotyping data on 713 trios to establish heritability of the 1q21.1 CNV. Families consisted of individuals with no reported incidence or history of cancer; however, other conditions were reported. The condition reported most frequently was autism (>50% of probands). Data from 565 trios was generated from lymphoblastoid cell lines and the remaining 148 were from peripheral blood. CNV profiles were generated for each offspring and parent in the trios using a family-based method in PennCNV to increase sensitivity of CNV detection[16]. The concordance frequency between offspring and parent was computed and reported as an estimated inheritance rate for the CNV.

Fluorescent *in situ* hybridization. Metaphase spreads were prepared either from peripheral blood lymphocytes or from subject-derived, lymphoblastoid cell lines using standard methodology. Chromosomes were visualized by counterstaining with 4,6-diamidino-2-phenylindole (DAPI). BAC and fosmid clones used for FISH analysis were obtained from BACPAC Resources. FISH was performed using fosmid W12-1967b11 (GenBank: G248P87625A6) and BAC RP4-790G17 (GenBank: AL138795). FISH analysis was carried out as described previously[31]. BACs and fosmids were isolated using the UltraClean plasmid kit (MoBio Labs Inc.) and probes were labelled with Spectrum Red or Green (Abbott Molecular Inc.) by nick translation. FISH images were captured using MacProbe software (Applied Imaging).

Realtime quantitative PCR validation of DNA copy number. Primers and probes were designed using Primer Express 3.0 (Applied Biosystems) with default parameters. Amplification primers were synthesized by IDT and probes were made by Applied Biosystems. Reactions were set up in triplicate using 10 ng of genomic DNA in a 10 μl reaction that contained 200 nM of probe, 900 nM of each amplification primer and 1× of Real-time PCR Master Mix (Applied Biosystems). Samples were amplified on an Applied Biosystems 7900HT Sequence Detection System using standard cycling conditions and data collected and analysed with SDS 2.3 software. Standard curves were constructed using serial twofold dilutions of genomic DNA from an individual without 1q21.1 copy number variation and used to estimate amounts of DNA in the experimental samples from their cycle threshold (C_t) values. Ratios of amounts were calculated from the assay designed in

the area of variation (1q21.1_DEL) and the assay in a region known to be present in all samples from the genome-wide association data (CTRL_2C). This normalized amount was then compared to the value in a control calibrator sample to produce a fold change ratio (normal = 1) and multiplied by 2 to generate a copy number (normal = 2). The following primers and probes were used: 1q21.1_DEL_Probe: 5′-6FAM-CACCACTGTCGTCCCTA-NFQ-3′; 1q21.1_DEL_ForwardPrimer: 5′-CCCTAAACATATGTGGGTGTACACA-3′; 1q21.1_DEL_ReversePrimer: 5′-TGCAGACAGACCCTATAGTGAGGTA-3′; CTRL_2C_Probe: 5′-6FAM-AAAAGGCACTGGTTAGGGA-MGB-NFQ-3′; CTRL_2C_FowardPrimer: 5′-CAAGTGCCAACAGAGTTGCTAGA-3′; CTRL_2C_ReversePrimer: 5′-TAATGAAGGAAGAGAATCAGTTCAGATT-3′.

PCR of cDNA and sequencing of 1q21.1 transcript. Standard PCR primers were designed using Primer3 with default parameters and synthesized by IDT as follows: left primer: 5′-CGTGCATTCATTTCCTTTGA-3′; right primer: 5′-GTGCACTGAATGGGGAAGTT-3′.

Complementary DNA was prepared from 4 μg of total RNA according to the SuperScript First-Strand Synthesis System for RT–PCR standard protocol using random hexamer primers (Invitrogen Life Technologies). PCR amplification of target cDNA was performed using the GC-RICH PCR system (Roche Applied Science) in a 50 μl reaction containing 5 μl cDNA, 0.2 mM dNTP mixture, 0.5 μM forward and reverse primers, 0.5 M GC-RICH Resolution Solution, 10 μl GC-RICH Reaction Buffer and 4 U GC-RICH Enzyme Mix. Amplification of the 1q21.1 transcript was performed in a Bio-Rad DNAEngine Peltier Thermal Cycler (Bio-Rad Laboratories) according to the following protocol: initial denaturation and enzyme activation at 95 °C for 8 min; denaturation at 95 °C for 1 min; primer annealing at 54 °C for 1 min; extension at 72 °C for 45 s for 40 total cycles; final extension at 72 °C for 10 min. PCR products were cloned into a pCR4-TOPO vector according to standard protocols (Invitrogen Life Technologies) and bi-directional sequencing of the cloned products was performed using M13 standard forward and reverse primers.

RNA samples for 'normal' panel. Fetal spinal ganglia were obtained at the Children's Hospital of Philadelphia from perinatal autopsies, in non-macerated fetuses ranging in gestational age from 18 weeks to 22 weeks who had undergone intrauterine demise. After removal of the lungs, the thoracic paraspinal sympathetic chains were identified *in situ* and entirely removed. Gestational age was recorded, but the specimens were otherwise anonymous. The specimens were frozen in optimum cutting temperature compound and stored at −70 °C. Validation of the presence of sympathetic ganglia was performed through examination of haematoxylin-and-eosin-stained frozen sections by a pathologist. Total RNA for all other normal adult and fetal tissues was obtained from Clontech or BioChain.

Real-time quantitative reverse transcriptase PCR. Primers and probes to assess mRNA expression in the samples were designed using Primer Express 3.0 (Applied Biosystems) with default parameters. Amplification primers were synthesized by IDT and probes were made by Applied Biosystems. Total RNA was DNase-treated and cDNA was prepared from 1 μg of total RNA according to the SuperScript First-Strand Synthesis System for RT–PCR standard protocol using random hexamer primers (Invitrogen Life Technologies) and diluted 1:4 for the unknowns. Reactions were set up in triplicate and also run with and without reverse transcriptase to confirm cDNA samples were free of genomic DNA contamination. Amplification was performed on an Applied Biosystems 7900HT Sequence Detection System using standard cycling conditions, and data collected and analysed with SDS 2.3 software. Standard curves were constructed using serial twofold dilutions of cDNA from either fetal brain or the neuroblastoma cell line IMR5. Ratios of mean quantities were normalized by comparing *NBPF23* expression to that of an endogenous control gene (*GAPDH* or *HPRT*). Primer and probe information is as follows: NBPF23_Probe: 5′-6FAM-TCCTCCTGAGGGTGTCT-NFQ-3′; NBPF23_ForwardPrimer: 5′-TCACCAGCTGATAGTCCCTTACC-3′; NBPF23_ReversePrimer: 5′-CCAAGAGCACACAGCACTGAA-3′.

26. Brodeur, G. M. *et al.* Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *J. Clin. Oncol.* **11**, 1466–1477 (1993).
27. Shimada, H. *et al.* The international neuroblastoma pathology classification (the Shimada system). *Cancer* **86**, 364–372 (1999).
28. Mathew, P. *et al.* Detection of *MYCN* gene amplification in neuroblastoma by fluorescence *in situ* hybridization: a pediatric oncology group study. *Neoplasia* **3**, 105–109 (2001).
29. Look, A. T. *et al.* Clinical relevance of tumor cell ploidy and *N*-myc gene amplification in childhood neuroblastoma. A pediatric oncology group study. *J. Clin. Oncol.* **9**, 581–591 (1991).
30. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genet.* **37**, 549–554 (2005).
31. Shaikh, T. H. *et al.* Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum. Mol. Genet.* **9**, 489–501 (2000).

LETTERS

Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*

Michael J. Wheeler[1]*†, Barend H. J. de Graaf[1]*†, Natalie Hadjiosif[1]*, Ruth M. Perry[1], Natalie S. Poulter[1], Kim Osman[1], Sabina Vatovec[1], Andrea Harper[1], F. Christopher H. Franklin[1] & Vernonica E. Franklin-Tong[1]

Higher plants produce seed through pollination, using specific interactions between pollen and pistil. Self-incompatibility is an important mechanism used in many species to prevent inbreeding; it is controlled by a multi-allelic *S* locus[1,2]. 'Self' (incompatible) pollen is discriminated from 'non-self' (compatible) pollen by interaction of pollen and pistil *S* locus components, and is subsequently inhibited. In *Papaver rhoeas*, the pistil *S* locus product is a small protein that interacts with incompatible pollen, triggering a Ca^{2+}-dependent signalling network, resulting in pollen inhibition and programmed cell death[3–7]. Here we have cloned three alleles of a highly polymorphic pollen-expressed gene, *PrpS* (*Papaver rhoeas* pollen *S*), from *Papaver* and provide evidence that this encodes the pollen *S* locus determinant. *PrpS* is a single-copy gene linked to the pistil *S* gene (currently called *S*, but referred to hereafter as *PrsS* for *Papaver rhoeas* stigma *S* determinant). Sequence analysis indicates that *PrsS* and *PrpS* are equally ancient and probably co-evolved. *PrpS* encodes a novel ~20-kDa protein. Consistent with predictions that it is a transmembrane protein, PrpS is associated with the plasma membrane. We show that a predicted extracellular loop segment of PrpS interacts with PrsS and, using *PrpS* antisense oligonucleotides, we demonstrate that PrpS is involved in *S*-specific inhibition of incompatible pollen. Identification of *PrpS* represents a major advance in our understanding of the *Papaver* self-incompatibility system. As a novel cell–cell recognition determinant it contributes to the available information concerning the origins and evolution of cell–cell recognition systems involved in discrimination between self and non-self, which also include histocompatibility systems in primitive chordates and vertebrates.

It has been established that self-incompatibility has evolved independently several times. Three self-incompatibility systems have been well characterized at a molecular level[1,2]. Both pollen and pistil *S* determinants are expected to have co-evolved and be physically linked to the *S* locus in order to maintain a functional self-incompatibility system. Other characteristics expected of them are high levels of allelic polymorphism and tissue-specific expression. Most importantly, they should function in mediating the self-incompatibility response. To understand fully how different self-incompatibility systems operate, identification of both the pistil and pollen *S* locus components, together with establishing mechanisms involved in pollen inhibition, is crucial. Previously, we identified the pistil *S* determinant for *Papaver rhoeas*[8–10] and established several components involved in pollen inhibition[3–7,11]. Although we identified a glycoprotein in pollen that bound to the pistil S protein, studies indicated that it was not the pollen *S* determinant, although it might modulate the self-incompatibility response[12]. Recent analysis of the *S₁* locus enabled identification of

the pollen component of the *S* locus on a cosmid clone comprising a 42-kilobase (kb) region at the *S₁* locus.

Nucleotide sequencing and analysis identified a novel putative open reading frame (ORF) 457 base pairs (bp) from the *S₁* pistil gene (Fig. 1a). Expression analysis using polymerase chain reaction with reverse transcription (RT–PCR) revealed that the ORF was specifically transcribed in pollen (Fig. 1b), appearing during anther development (Fig. 1c). The temporal expression pattern is very similar to that of the pistil *S* gene[8]. These data suggested that this ORF was a candidate for the *Papaver* pollen *S* gene (designated *PrpS* for *Papaver rhoeas* pollen *S*). We propose renaming the gene that determines self-incompatibility specificity in pistils (currently designated as *S*) to provide a clearer nomenclature; we suggest *PrsS* (*Papaver rhoeas* stigma *S* determinant).

The cDNA of *PrpS₁* comprises 1,206 bp containing a coding region of 579 bp encoding a 192-amino-acid polypeptide with a predicted molecular mass of 20.5 kDa, pI 7.55. We subsequently cloned *PrpS₃* and *PrpS₈* from *S₃S₈* pollen RNA. The *PrpS₃* and *PrpS₈* coding sequences are 576 bp and 582 bp (191 and 193 amino acids), respectively (Fig. 1d); *PrpS₃* and *PrpS₈* encode proteins of predicted molecular mass of 21.1 kDa (pI 6.57) and 20.9 kDa (pI 8.51), respectively. Southern blotting revealed that *PrpS* is a single-copy gene (Supplementary Fig. 1), so the related sequences identified as *PrpS₃* and *PrpS₈* are clearly allelic to *PrpS₁*, rather than being related/paralogous genes.

Segregation analyses were conducted to obtain evidence of genetic linkage at the *S* locus between *PrsS₃* and *PrpS₃* and between *PrsS₈* and *PrpS₈*. Specific primers were used to amplify regions of the pistil *PrsS₁*, *PrsS₃* and *PrsS₈* and pollen *PrpS₁*, *PrpS₃* and *PrpS₈* genes from genomic DNA from two full-sibling families segregating for these haplotypes (45 and 25 individuals, a total of 140 *PrpS*/*PrsS* pairs). *PrpS₁* was amplified only from plants carrying the *S₁* haplotype; *PrpS₃* was amplified only from plants carrying the *S₃* haplotype; and *PrpS₈* was amplified only from plants carrying the *S₈* haplotype. The pistil *PrsS₁*, *PrsS₃* and *PrsS₈* sequences were also amplified only from plants carrying the respective *S* haplotypes (Fig. 1e), as expected. This demonstrates co-segregation and linkage of *PrpS₁*, *PrpS₃* and *PrpS₈* and their cognate *PrsS* genes, as no recombination was detected (recombination frequency <0.021). Thus, at the 95% rejection level we can be confident that there is no recombination.

A markedly high level of allelic sequence polymorphism is a well-documented feature of *S* locus proteins; *S* alleles have unusually high amino acid sequence divergence within species[13–16]. *Papaver* is no exception, with the pistil proteins PrsS₁ and PrsS₃ showing 46% sequence divergence, PrsS₁ and PrsS₈ showing 40% divergence, and PrsS₃ and PrsS₈ with 46% divergence. PrpS proteins have a similar
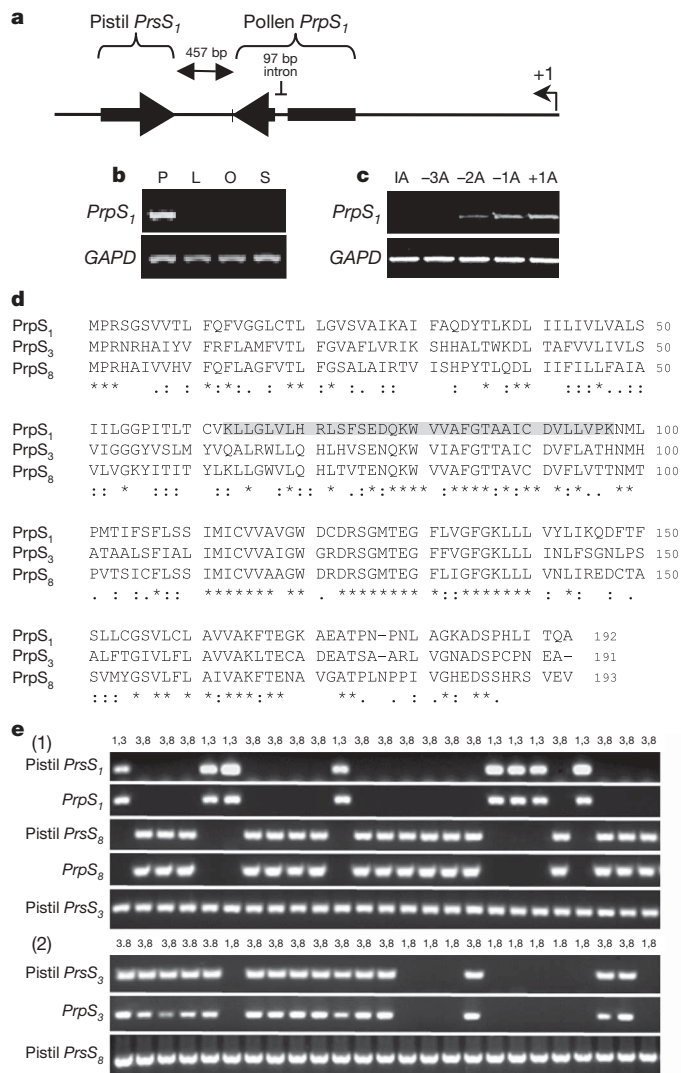
Figure 1 | Organization and expression of PrpS₁. a, Organization of the S_1 locus. Arrows indicate pistil *PrsS₁* and pollen *PrpS₁* coding sequences and their orientation. Transcription start site (+1) is shown. An intron is located 84 bp from the 3′ end. **b**, RT–PCR shows that *PrpS₁* is expressed in pollen (P) but not in leaf (L), ovary (O) or stigma (S). **c**, RT–PCR showing that *PrpS₁* expression increases during anther development. IA, immature anthers; −3A, −2A, −1A, anthers 3, 2 and 1 days pre-anthesis, respectively; +1A, anthesis. Glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) shows equal loading. **d**, Alignment of PrpS₁, PrpS₃ and PrpS₈ deduced amino acid sequences. The predicted extracellular loop segment (TMHMM) is indicated for PrpS₁ (grey box). '*', identical; ':', conserved, '.', semi-conserved amino acid substitutions. **e**, Linkage of *PrpS* and *PrsS* to the *S* locus. Full-sibling families segregating for haplotypes (1) S_1S_3 and S_3S_8 and (2) S_1S_8 and S_3S_8 were used for PCR. Pistil *PrsS₁*, *PrsS₃* and *PrsS₈*, and pollen *PrpS₁*, *PrpS₃* and *PrpS₈* sequences were amplified only if plants carried the corresponding *PrpS* allele. *S* haplotypes are indicated: S_1S_3 (1,3), S_3S_8 (3,8), S_1S_8 (1,8).

level of polymorphism (Fig. 1d): the PrpS₁ and PrpS₃ predicted amino acid sequences are 50% divergent; PrpS₁ and PrpS₈ show 40% divergence; and PrpS₃ and PrpS₈ are 47% divergent.

The pollen and pistil *S* determinants should show evidence of co-evolution[16]. Examination of the *PrpS* sequences for non-synonymous to synonymous (K_a/K_s) substitutions reveal that the *PrsS* alleles have a mean K_a/K_s ratio of 0.234, and the *PrpS* alleles have a mean K_a/K_s ratio of 0.368 (Supplementary Table 1). A two-tailed *t*-test showed no significant difference between substitution rates in *PrpS* and *PrsS* genes. These data indicate that the pollen and pistil *S* alleles co-evolved and are likely to be similarly ancient.

PrpS has no significant sequence homology to any protein in existing databases. Sequence analysis, using a range of prediction programs,

indicated that PrpS has 3–5 predicted transmembrane helices, and alignment of the three *PrpS* alleles indicates that they share a similar topology (Fig. 1d and Supplementary Fig. 2). In support of predictions that PrpS is a transmembrane protein, western blot analysis using antisera raised against PrpS₁ revealed that PrpS₁ was detected as a ~20-kDa protein specifically in S_1 pollen-membrane-enriched extracts (Fig. 2a–c). Moreover, immunolocalization studies revealed that PrpS₁ is associated with the pollen tube plasma membrane (Fig. 2d). Although the PrpS sequences do not show any particular bias for the 'positive inside rule'[17], structural predictions suggest an extracellular loop segment, comprising amino acids ~60–100 (63–97 using TMHMM, Fig. 1d; see also Supplementary Fig. 2). We hypothesized that this region might be involved in the interaction with PrsS and show that a peptide corresponding to part of the predicted PrpS extracellular loop interacted with the PrsS protein, whereas the corresponding randomized peptide did not (Fig. 2e).

To determine whether PrpS is functionally involved in the self-incompatibility response, we investigated whether it mediates *S*-specific pollen inhibition, using *in vitro* self-incompatibility bioassays[8]. Peptides based on extracellular domains of receptors have been used to identify ligand-binding epitopes via their ability to block the receptor–ligand interaction[18]. Preliminary experiments with the peptide used in the binding assay tested whether it could block self-incompatibility-mediated inhibition. Pollen from plants with haplotypes S_1S_3, when challenged with incompatible recombinant PrsS, were rescued from inhibition by PrpS₁ peptides ($n = 6$; $P < 0.001$; Supplementary Fig. 3 and Supplementary Table 2) whereas randomized peptides based on the same amino acids had no effect ($n = 6$; not significant, Supplementary Fig. 3 and Supplementary Table 2). This is consistent with the hypothesis that this region is involved in recognition and indicated that PrpS might mediate pollen inhibition. To confirm this possibility, we used an antisense oligonucleotide approach[3,19]. We hypothesized that if PrpS functions as the pollen *S* determinant, knockdown of its expression should result in alleviation of pollen tube inhibition in an *S*-specific manner. We induced self-incompatibility *in vitro* in the presence of either antisense (as-ODNs) or sense (s-ODNs) oligonucleotides to test this hypothesis. As our plants are heterozygous for *S* haplotypes, the pollen phenotype of
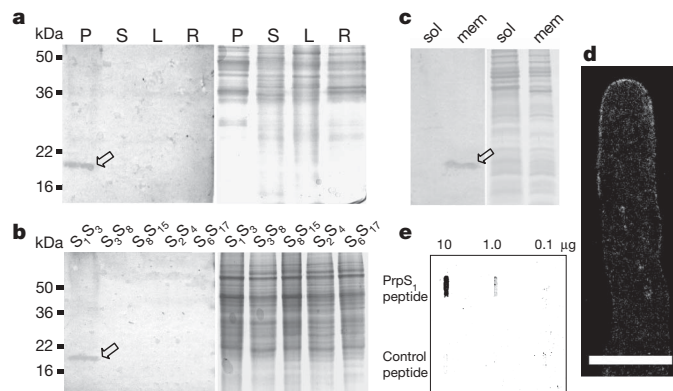

Figure 2 | PrpS is pollen-membrane associated. a, Western blot detects PrpS₁ at ~20 kDa (arrow) in pollen (P), but not in stigma (S), leaf (L) or root (R) membrane-enriched protein extracts (left panel). Coomassie staining shows equal loading (right panel). **b**, *PrpS₁* is expressed in pollen samples carrying S_1, but not other alleles (left panel). Coomassie staining shows equal loading (right panel). **c**, Western blot of fractionated pollen extracts. PrpS is not present in cytosolic extracts (sol), but is present in a Triton-X-100 enriched fraction (mem). **d**, Immunolocalization shows PrpS₁ localization to the pollen tube plasma membrane. Scale bar, 10 μm. **e**, PrpS₁ binds PrsS₁. A 15-mer peptide corresponding to part of the PrpS₁ predicted 35 amino acid extracellular loop region (DQKWVVAFGTAAICD) binds recombinant PrsS₁ in a concentration-dependent manner (top). A corresponding randomized peptide (FTVDVKDCAAAWGQI) did not bind PrsS₁ (bottom). Concentrations are as indicated; see Methods for details; $n = 8$.

plants with S_1S_3 haplotypes should theoretically be 50% S_1 and 50% S_3. Thus, if the interaction is S specific, as-ODNs specific for $PrpS_1$ should only affect 50% of pollen (carrying S_1).

Self-incompatibility induced strong inhibition of pollen tube growth (a 79% reduction in length compared with the controls) and we observed a significant alleviation of this inhibition in an incompatible combination in the presence of as-ODNs but not with corresponding s-ODNs (Fig. 3). With pollen from plants with S_1S_3 haplotypes, self-incompatibility induced strong inhibition of pollen tube length (22.1%, $P < 0.001$, $n = 300$); addition of $PrpS_1$ as-ODNs gave a highly significant recovery of self-incompatibility-treated tubes (58.3% increase in length compared to self-incompatibility-treated; $P < 0.001$, $n = 150$), and incompatible pollen responded in a bimodal manner, consistent with only S_1 pollen being affected (Supplementary Fig. 4). When $PrpS_8$ as-ODNs were added to the same pollen from plants with S_1S_3 haplotypes, they did not alleviate self-incompatibility-induced inhibition ($P = 0.604$, not significant, $n = 150$). This demonstrates that the $PrpS_1$ and $PrpS_8$ as-ODNs had an S-specific effect. As expected, $PrpS_1$ s-ODNs did not affect the self-incompatibility response ($P = 0.591$, not significant, $n = 150$).

To confirm further the S-specific effect of the as-ODNs, we also tested their effect on pollen from plants with haplotypes S_3S_8. Self-incompatibility resulted in inhibited pollen tubes (19.8% of the control, $n = 300$), and addition of $PrpS_8$ as-ODNs alleviated the self-incompatibility-induced inhibition, giving a highly significant 100.3% increase in pollen tube length ($P < 0.001$, $n = 150$), whereas there was no effect using $PrpS_1$ as-ODNs ($P = 0.336$, not significant, $n = 150$) or $PrpS_8$ s-ODNs ($P = 0.565$, not significant, $n = 150$) (Fig. 3). Together these data demonstrate that PrpS has a crucial role in self-incompatibility-induced S-haplotype-specific pollen tube inhibition.

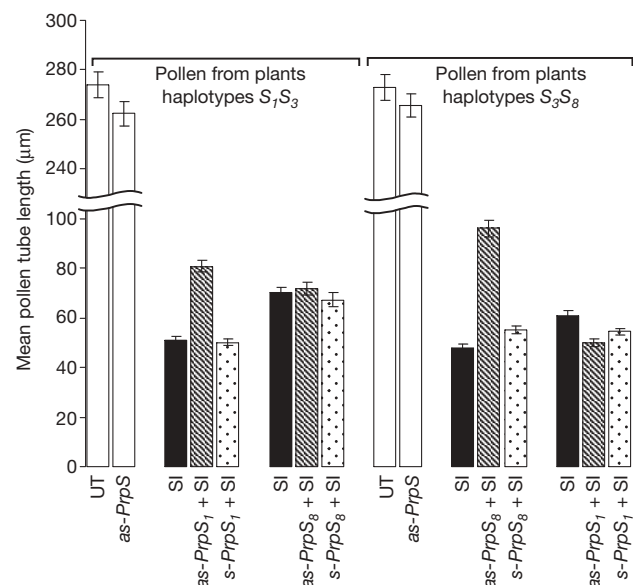We have cloned a polymorphic pollen-expressed gene, $PrpS$. Together our data are consistent with the hypothesis that PrpS is the *Papaver* pollen S determinant as it mediates S-specific recognition and inhibition. Self-incompatibility in *Papaver* is distinct from the other well characterized self-incompatibility systems (that is, from both the *Brassica* pollen S determinant SCR/SP11 and the pollen F-box protein SLF/SFB from the S-RNase-based self-incompatibility system)[1,20–23]. As $PrpS$ has no homologues, its nature is intriguing. Self/non-self discrimination and other recognition systems which are controlled by a highly polymorphic locus are not limited to self-incompatibility; other systems include disease resistance in plants[24] and histocompatibility systems in animals[25–27]. These parallels between non-analogous recognition systems were recognized, and their importance appreciated, long before the molecular basis of these systems were elucidated[25] and the nature of their polymorphism has intrigued population and evolutionary biologists for decades. The identification of PrpS as a novel cell–cell recognition determinant thus contributes to the available information regarding the evolution of self/non-self recognition systems.

METHODS SUMMARY

Cloning of $PrpS_1$, $PrpS_3$ and $PrpS_8$. A genomic clone of $PrpS_1$ was identified by nucleotide sequence analysis of a 42-kb clone carrying the S_1 locus, obtained by screening a *P. rhoeas* S_1S_3 cosmid genomic DNA library (SuperCos1, Stratagene) with $PrsS_1$ cDNA. The DNA upstream and downstream of the $PrsS_1$ gene was sequenced and analysed using BLAST (http://ncbi.nlm.nih.gov/BLAST) and ORF Finder (http://searchlauncher.bcm.tmc.edu)[28]. The organization of $PrpS_1$ and $PrsS_1$ genes was confirmed using PCR on genomic DNA of S_1- and non-S_1-containing plants. $PrpS_3$ and $PrpS_8$ cDNAs were obtained using RT–PCR, 3′ and 5′-RACE PCR (see Methods for primer details) on pollen cDNA from suitable S-haplotypes, using low annealing temperatures (48 °C).

K_a/K_s calculations. DNAsP[29] was used to estimate K_a (the number of non-synonymous substitutions per non-synonymous site) and K_s (the number of synonymous substitutions per synonymous site) for pairs of $PrsS$ and $PrpS$ nucleotide sequences.

Peptide binding assay. A 15-amino-acid peptide (DQKWVVAFGTAAICD) corresponding to part of the predicted extracellular loop segment of PrpS$_1$ (TMHMM; http://www.cbs.dtu.dk/services/TMHMM30) and a randomized version (FTVDVKDCAAAWGQI) were synthesized (Alta Bioscience). The peptides (10 μg, 1 μg, 0.1 μg) were bound to PVDF membrane. This was incubated with recombinant PrsS$_1$ and then probed for binding using anti-PrsS$_1$ antisera and alkaline phosphatase detection.

Antisense oligonucleotide silencing of $PrpS$ expression. Phosphorothioated gene-specific antisense oligodeoxynucleotides (as-ODN) and their sense controls (s-ODN) were designed ($PrpS_1$ as-ODN, gtccTCCCAGTATTAttga; $PrpS_1$ s-ODN, tcaaTAATACTGGGAggac; $PrpS_8$ as-ODN, ttccCACCAGCACAGCaatt; $PrpS_8$ s-ODN, aattGCTGTGCTGGTGggaa; lowercase letters indicate bases linked by phosphorothioate bonds). Pollen was grown *in vitro* and pre-treated with as-ODNs and s-ODNs[3,19] for 1 h before induction of self-incompatibility with recombinant PrsS$_1$, PrsS$_3$ and PrsS$_8$ (ref. 8). After 2 h, pollen tubes were fixed in 2% formaldehyde and 150 pollen tube lengths were measured in three independent experiments.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489 (2005).
2. Franklin-Tong, V. E. (ed.) *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms* (Springer, 2008).
3. de Graaf, B. H. J. *et al.* Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. *Nature* **444**, 490–493 (2006).
4. Franklin-Tong, V. E., Ride, J. P., Read, N. D., Trewavas, A. J. & Franklin, F. C. H. The self-incompatibility response in *Papaver rhoeas* is mediated by cytosolic-free calcium. *Plant J.* **4**, 163–177 (1993).
5. Snowman, B. N., Kovar, D. R., Shevchenko, G., Franklin-Tong, V. E. & Staiger, C. J. Signal-mediated depolymerization of actin in pollen during the self-incompatibility response. *Plant Cell* **14**, 2613–2626 (2002).
6. Thomas, S. G. & Franklin-Tong, V. E. Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* **429**, 305–309 (2004).
7. Bosch, M. & Franklin-Tong, V. E. Temporal and spatial activation of caspase-like enzymes induced by self-incompatibility in *Papaver* pollen. *Proc. Natl Acad. Sci. USA* **104**, 18327–18332 (2007).

Figure 3 | PrpS determines S-specific pollen inhibition. $PrpS_1$ and $PrpS_8$ antisense oligonucleotides (as-ODNs: as-$PrpS_1$, as-$PrpS_8$) 'rescue' pollen from plants with S_1S_3 or S_3S_8 haplotypes from self-incompatibility (SI)-induced inhibition in an S-specific manner, whereas $PrpS_1$ and $PrpS_8$ sense oligonucleotides (s-ODNs: s-$PrpS_1$, s-$PrpS_8$) do not. Controls: untreated (UT) pollen and as-ODNs without self-incompatibility induction (white bars); self-incompatibility-induced pollen (black bars); self-incompatibility-induced in the presence of as-ODN (crosshatched bars); self-incompatibility-induced in the presence of s-ODNs (dotted bars). Fifty pollen tubes were measured in three independent experiments (150 in total); error bars indicate s.e.m.

8. Foote, H. C. C. *et al.* Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. *Proc. Natl Acad. Sci. USA* **91**, 2265–2269 (1994).

9. Kurup, S. *et al.* Identification and cloning of related self-incompatibility *S*-genes in *Papaver rhoeas* and *Papaver nudicaule. Sex. Plant Reprod.* **11**, 192–198 (1998).

10. Walker, E. A. *et al.* Molecular analysis of two functional homologues of the S_3 allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. *Plant Mol. Biol.* **30**, 983–994 (1996).

11. Thomas, S. G., Huang, S., Li, S., Staiger, C. J. & Franklin-Tong, V. E. Actin depolymerization is sufficient to induce programmed cell death in self-incompatible pollen. *J. Cell Biol.* **174**, 221–229 (2006).

12. Hearn, M. J., Franklin, F. C. H. & Ride, J. P. Identification of a membrane glycoprotein in pollen of *Papaver rhoeas* which binds stigmatic self-incompatibility (*S*-) proteins. *Plant J.* **9**, 467–475 (1996).

13. Ioerger, T. R., Clark, A. G. & Kao, T.-H. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc. Natl Acad. Sci. USA* **87**, 9732–9735 (1990).

14. Kohn, J. R. in *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms* (ed. Franklin-Tong, V. E.) 103–121 (Springer, 2008).

15. Charlesworth, D. Multi-allelic self-incompatibility polymorphisms in plants. *Bioessays* **17**, 31–38 (1995).

16. Newbigin, E., Paape, T. & Kohn, J. R. RNase-based self-incompatibility: puzzled by pollen S. *Plant Cell* **20**, 2286–2292 (2008).

17. von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678 (1988).

18. Biris, N. *et al.* Mapping the binding domains of the αIIb subunit. *Eur. J. Biochem.* **270**, 3760–3767 (2003).

19. Moutinho, A. *et al.* Antisense perturbation of protein function in living pollen tubes. *Sex. Plant Reprod.* **14**, 101–104 (2001).

20. McClure, B. & Franklin-Tong, V. Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta* **224**, 233–245 (2006).

21. Sijacic, P. *et al.* Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**, 302–305 (2004).

22. Qiao, H. *et al.* The F-box protein AhSLF-S_2 controls the pollen function of S-RNase-based self-incompatibility. *Plant Cell* **16**, 2307–2322 (2004).

23. Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E. & Nasrallah, J. B. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea. Proc. Natl Acad. Sci. USA* **88**, 8816–8820 (1991).

24. Dangl, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001).

25. Burnet, F. M. "Self-recognition" in colonial marine forms and flowering plants in relation to the evolution of immunity. *Nature* **232**, 230–235 (1971).

26. De Tomaso, A. W. *et al.* Isolation and characterization of a protochordate histocompatibility locus. *Nature* **438**, 454–459 (2005).

27. Scofield, V. L., Schlumpberger, J. M., West, L. A. & Weissman, I. L. Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* **295**, 499–502 (1982).

28. Worley, K. C., Wiese, B. A. & Smith, R. F. BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Res.* **5**, 173–184 (1995).

29. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., Rozas, R. & Dna, S. P. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497 (2003).

30. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

METHODS

Sequence homology comparisons. Sequence identities were calculated for PrsS and PrpS, using pairwise comparisons of the amino acids comprising the mature peptide, using BLAST (BLOSUM62)[31].

Southern blotting. Southern blots of *Papaver* genomic DNA from plants of different *S*-haplotypes (S_1S_8, S_1S_3, S_3S_8, S_2S_4, S_3S_4, S_7S_x) digested with EcoRV, BamHI, Pst1 or BglII were probed with a *PrpS_1* probe. Blots were hybridized at 60 °C overnight and washed down to ×0.5 SSC at 50 °C, and signals detected using autoradiography.

Structural predictions for PrpS and interpretation of the data. We analysed PrpS sequences using a number of protein prediction programs: TMHMM2.0[32], PredictProtein, SOSUI, HMMTOP, TMpred, TM-Finder, SPLIT 4, ConPredII, Phobius. TMHMM2.0 in particular is regarded as a very robust transmembrane helix predictor, although predicting the number of membrane-spanning regions remains difficult[33]. Although predictions differ, they all predict that PrpS has transmembrane helices. We have been advised that the predictions suggest anything between 3–5 transmembrane segments and the alignment of the PrpS sequences suggests that all three proteins share a similar topology. Although many predictions indicated three or five transmembrane domains, it is more likely that PrpS has four transmembrane segments, as it is a good number to make a four-helix bundle in the membrane, as three and five transmembrane proteins are rare (A. Lovering, personal communication). All the predictions indicate an extracellular loop segment. Supplementary Fig. 2 shows a cartoon indicating predictions for the regions; this is a tentative assignment of possible topology.

Linkage analysis. Segregation analysis was carried out on individual plants from several full-sibling families which share three *S* haplotypes and segregate for haplotypes S_1S_3 or S_3S_8 and S_1S_8 or S_3S_8. These are well established families, in which *S* haplotypes were designated ~20 years ago, produced using controlled crosses between individuals which had their *S*-haplotypes verified using aniline blue microscopy; they have been analysed for *S* allele segregation for at least seven generations in a pedigree going back to 1994.

The number of plants required to be analysed, in order to be statistically confident at 95% rejection level[34] that there is no recombination, is 25 plants (one cannot formally provide evidence that there is never recombination). We carried out analysis of a full-sibling family, segregating for pollen *PrpS_1* and *PrpS_3* and pistil *PrsS_1* and *PrsS_3* alleles (25 plants, 50 alleles) and for the S_1 and S_8 loci, and analysed 45 full-sibling plants (90 alleles) to show segregation of pollen *PrpS_1* and *PrpS_8* and pistil *PrsS_1* and *PrsS_8*. Thus, we examined segregation of a total of 140 *PrpS/PrsS* pairs of alleles. Genomic DNA from leaf tissue was extracted (Extract-N-Amp Plant PCR kit, Sigma-Aldrich) and gene-specific primers used to specifically amplify *PrsS_1*, *PrsS_3* and *PrsS_8*, and *PrpS_1*, *PrpS_3* and *PrpS_8* sequences. (See Supplementary Table 3 for primer details.) No recombination was detected (RF < 0.021). Thus, at the 95% rejection level, we can be confident that there is no recombination.

RT–PCR to show tissue and developmental specificity. Standard RT–PCR techniques were used for expression analysis. Total RNA was extracted from anthers from plants with the S_1 haplotype at different stages of development and from different tissues (RNAeasy plant mini kit, QIAGEN) and cDNA synthesized (Omniscript RT kit, QIAGEN). Gene-specific primers (see Supplementary Table 3) were used to amplify *PrpS_1* transcripts; primers for the *P. rhoeas* glyceraldehyde-3-phosphate dehydrogenase (*GAPD*) gene acted as controls.

Production of antisera. The predicted 60 amino acid carboxy terminus of *PrpS_1* (designated PrpS1-60C) was expressed as a His-tagged recombinant protein using pET21b (Novagen). Recombinant protein was isolated from *Escherichia coli* BL21 (DE3) using Ni-NTA resin following the manufacturer's (QIAGEN) protocol. Antisera (PrpS1-60C) was raised in rats (ISL Immune Systems).

Protein extraction for SDS–PAGE and western blotting. Extracts enriched for membrane proteins were made in 100 mM Tris-HCl pH 8, 200 mM NaCl, 2 mM EDTA, 1 M sucrose, 0.5% Triton X-100, Protease Inhibitor cocktail (Roche). Protein concentrations were determined[35] and proteins separated using SDS–PAGE and electroblotted (400 mA, 3 h) onto Hybond C membranes (GE Healthcare). These were incubated with the PrpS1-60C antibody (1:2,000) for 2 h, followed by alkaline-phosphatase-conjugated anti-rat secondary antibody (Sigma); detection used BCIP (5-bromo-4-chloro-3′-indolyphosphate p-toluidine salt) and NBT (Nitroblue tetrazolium chloride).

Immunolocalization. *Papaver rhoeas* pollen from plants with haplotypes S_1S_3 was grown in germination medium (GM)[36] for 1 h at 25 °C and prepared according to the method of ref. 37, followed by incubation with the PrpS1-60C antibody (1:500 in TBS + 1% BSA; 4 °C overnight), then FITC-conjugated goat anti-rat antibody (1:50, 1.5 h). Cells were mounted on slides and examined by confocal microscopy (Bio-Rad Radiance 2000 MP) using single scans (×100 plan-Apo 1.4 NA oil objective, Nikon). When pre-immune antiserum was used, no signal was obtained using identical additions and settings.

The *in vitro* bioassay for *Papaver* self-incompatibility: use of this as an assay for function. We have used *in vitro* bioassays for demonstration of PrpS function instead of the alternative strategy of stable transformation with a *PrpS* allele because the latter is not possible in *Papaver*. These bioassays have been routinely used to demonstrate *S*-specific function. Because it has not been possible to generate plants that are homozygous for *S* haplotypes in *Papaver*, we use full-sibling families that are segregating for plants that are heterozygous for *S* haplotypes. This makes the interpretation of the data slightly more complicated, as we always have a mixed population of pollen grains in our samples.

The *Papaver* self-incompatibility *in vitro* bioassay was initially used for demonstrating that stigmatic extracts had *S*-specific biological activity; we showed the distribution of individual pollen tube lengths for control, fully compatible, incompatible and half-compatible interactions, with bimodal distribution of inhibition in a half-compatible self-incompatibility response, as expected[38]. This self-incompatibility bioassay was used to provide evidence for the pistil *S* gene (for which we propose the name *PrsS*) being the *S* determinant[39]; we demonstrated *S*-allele-specific inhibition of pollen by recombinant PrsS_1 protein. This half-compatible interaction gave 40–44% mean inhibition (compared to the theoretical maximum of 50%) with pollen from plants with haplotypes S_1S_3 or S_1S_6, but no inhibition with pollen from plants with haplotypes S_3S_6 or S_2S_4. As we now have several cloned pistil *PrsS* alleles, for any *in vitro* self-incompatibility bioassay, we routinely add two recombinant PrsS proteins to obtain a full self-incompatibility response where all pollen is inhibited (see for example refs 40–42).

For the self-incompatibility *in vitro* bioassays carried out here, we used two recombinant PrsS proteins (PrsS_1 and PrsS_3 to pollen from plants with haplotypes S_1S_3, and PrsS_3 and PrsS_8 proteins to pollen from plants with haplotypes S_3S_8) to achieve a full self-incompatibility response (all pollen inhibited). With the peptide bioassay (see below) we added the PrpS_1 peptide to the self-incompatibility bioassay, so the expectation was that a maximum of 50% of pollen would be rescued if we obtained *S*-haplotype-specific alleviation of self-incompatibility. For the antisense oligonucleotide experiments we used pollen from plants with S_1S_3 and S_3S_8 haplotypes in combination with antisense *PrpS_1* and/or antisense *PrpS_8*. The expectation was that *S*-haplotype-specific alleviation of self-incompatibility would be obtained using antisense *PrpS_1* with pollen from plants with S_1S_3 haplotypes, but this would only rescue 50% (pollen carrying *PrpS_1*, but not pollen carrying *PrpS_3*); we would also expect no alleviation of self-incompatibility with antisense *PrpS_1* with the pollen from S_3S_8 plants. We also carried out experiments with antisense *PrpS_8* in these combinations, with the prediction that antisense *PrpS_8* will alleviate the self-incompatibility response by a maximum of 50% with pollen from plants with S_3S_8 haplotypes (rescue of pollen carrying *PrpS_8*, but not *PrpS_3*). Supplementary Fig. 3 shows data for populations of individual pollen tubes for an experiment using a *PrpS_1* antisense oligonucleotide added to a self-incompatibility assay using pollen from plants with S_1S_3 or S_3S_8 haplotypes. The expectation is that *PrpS_1* antisense oligonucleotide should rescue pollen carrying the *PrpS_1* allele and not the *PrpS_3* or *PrpS_8* alleles. The plot shows that the rescue is clearly bimodal, with some pollen tubes still strongly inhibited, whereas others have alleviated inhibition, which gives confidence to the mean data shown in Fig. 3.

We designed these experiments to be reciprocal so that we could demonstrate allelic specificity of rescue and to rule out some nonspecific effects, as if the antisense *PrpS_1* caused pollen tubes to grow longer for some other reason, we should see this effect with the pollen from plants with S_3S_8 haplotypes. Sense oligonucleotides were used as additional controls, in order to demonstrate that the antisense oligonucleotides specifically had this effect on pollen tube growth. These assays provide a robust test of whether *PrpS_1* and *PrpS_8* are allelic, as we show that they have the same biological function and they exhibit allelic specificity. If *PrpS_8* was not allelic to *PrpS_1*, then it would not affect pollen tube growth as predicted. Statistical tests were carried out using MINITAB. Tests comprised two-way comparisons between pairs of data using a Student's *t*-test.

Peptide bioassay. A 15-amino-acid peptide corresponding to part of the predicted 35-amino-acid PrpS_1 external loop region was designed and two randomized versions of this peptide were synthesized (see Methods Summary for details). *P. rhoeas* pollen was grown *in vitro* and self-incompatibility induced using recombinant PrsS_1, PrsS_3 and PrsS_8 (ref. 42). Self-incompatibility-induced pollen inhibition was compared with self-incompatibility in the presence of the PrpS_1 peptide or a randomized peptide at 200 μg ml^{-1}. Pollen grains and tubes were scored after 1 h, using two categories: 'inhibition' or 'growth'; a minimum of 100 pollen grains/tubes was scored for each sample. Data were analysed using Fisher's exact test for 2 × 2 contingency tables[43].

Primers used for the isolation of *PrpS* alleles. Gene-specific primers used to amplify full-length *PrpS_1*: PRPS1-5′-UTR1 GTAGCATTTACAATCTTCTTA GAAATGC; PRPS1-3′-UTR1 GAGAACACGTCATTGGAATTATTGAG.

5′ RACE primer to obtain another *S* allele (resulting in isolation of *PrpS_8*): NH3′-3 GCGACCGAAGTGGCATG.

Gene-specific primers used to amplify *PrpS₈*: PRPS8 5′-1 GGCAGTTATG CCTCGACATGCAATTG; PRPS8 3′-1 TTAAACCTCAACACTACGGTGG.

5′ RACE primer used to obtain full-length *PrpS₈*: PRPS8 5′RACE-1 GCTGTGCAATCCTCTCTGATCAAG.

Degenerate primers used to obtain *PrpS₃*: PRPS 5′-1 ATGCCACGAMR TGSAAKTGTTG; PRPS 5′-2 CCTATTKGGAKYCKCASTTGCC; PRPS 5′-3 GT AGTMGCATTTGGGACYRCTGC; PRPS 3′-1 GTGAACTTAGCAACAA YWGCRAGG; PRPS 3′-2 TCAAGTWKACTAGTARRAGCTTGCC; PRPS 3′-3 CCAAYYAAAAATCCYTCRGTCATGCC.

3′ and 5′ RACE primers used to obtain full-length *PrpS₃*: BGPRPS3-3′R1 GCTTCTACTGATTAACTTGTTCTCCGG; BGPRPS3-5′R1 CTGCAGTTGC ATGCATATTGTGTGTCG.

Gene-specific primers used for *S*-locus linkage analysis. Primers for pistil *PrsS₁*, *PrsS₃* and *PrsS₈*: SS1-5′ GCTATCGTTCTTCTTGTCCAAGTCAAGC GG; SS1-3′ CATCCCTCTTTGCCTGATAGGAATAAAACCCG; SS3-5′ GACTT TGGTTAGCATGTCCAATTCCATCGGC; SS3-3′ GTCCCTCTTGCCCGAGT AAGCATCG; SS8-5′ GTCCTTCTTGACCTTGGCCTCATCTCG; SS8-3′ CCGT GATCATCTCGTTGTGCTCGATAGG.

Primers for *PrpS₁*, *PrpS₃* and *PrpS₈*: PRPS1-5′ CAGTTTGTTGGAGGATT ATGTACCCCGTTGG; PRPS1-3′ GCGACCACACAGATCATTATGGAAGAT AAGAAGG; PRPS3-5′ GAGTAGCATTTCTCGTGAGAATTAAATCTCACCA TGC; PRPS3-3′ GGTCACGACCCCAACCAATTGCAACG; PRPS8-5′ CGCAC TTGCCATAAGAACTGTCATTTCTCACC; PRPS8-3′ CACTACGATCACGGT CCCAACCAGC.

31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

32. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

33. Möller, S., Croning, M. D. R. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653 (2001).

34. Fisher, R. A. *Statistical Methods and Scientific Inference* (Hafner, 1956).

35. Bradford, M. M. A dye binding assay for protein. *Anal. Biochem.* **72**, 248–254 (1976).

36. Snowman, B. N., Kovar, D. R., Shevchenko, G., Franklin-Tong, V. E. & Staiger, C. J. Signal-mediated depolymerization of actin in pollen during the self-incompatibility response. *Plant Cell* **14**, 2613–2626 (2002).

37. Poulter, N. S., Vatovec, S. & Franklin-Tong, V. E. Microtubules are a target for self-incompatibility signaling in *Papaver* pollen. *Plant Physiol.* **146**, 1358–1367 (2008).

38. Franklin-Tong, V. E., Lawrence, M. J. & Franklin, F. C. H. Self-incompatibility in *Papaver rhoeas*. 1. Characterization of the stigmatic component. *Heredity* **61**, 286 (1988).

39. Foote, H. C. C. *et al.* Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl Acad. Sci. USA* **91**, 2265–2269 (1994).

40. Geitmann, A., Snowman, B. N., Emons, A. M. C. & Franklin-Tong, V. E. Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*. *Plant Cell* **12**, 1239–1251 (2000).

41. Jordan, N. D., Franklin, F. C. H. & Franklin-Tong, V. E. Evidence for DNA fragmentation triggered in the self-incompatibility response in pollen of *Papaver rhoeas*. *Plant J.* **23**, 471–479 (2000).

42. Thomas, S. G. & Franklin-Tong, V. E. Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* **429**, 305–309 (2004).

43. Bailey, N. T. J. *Statistical Methods in Biology* (The English Universities Press, 1959).

LETTERS

A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish

Philipp Niethammer[1]*, Clemens Grabher[2]*†, A. Thomas Look[2,3] & Timothy J. Mitchison[1]

Barrier structures (for example, epithelia around tissues and plasma membranes around cells) are required for internal homeostasis and protection from pathogens. Wound detection and healing represent a dormant morphogenetic program that can be rapidly executed to restore barrier integrity and tissue homeostasis. In animals, initial steps include recruitment of leukocytes to the site of injury across distances of hundreds of micrometres within minutes of wounding. The spatial signals that direct this immediate tissue response are unknown. Owing to their fast diffusion and versatile biological activities, reactive oxygen species, including hydrogen peroxide (H_2O_2), are interesting candidates for wound-to-leukocyte signalling. Here we probe the role of H_2O_2 during the early events of wound responses in zebrafish larvae expressing a genetically encoded H_2O_2 sensor[1]. This reporter revealed a sustained rise in H_2O_2 concentration at the wound margin, starting ~3 min after wounding and peaking at ~20 min, which extended ~100–200 μm into the tail-fin epithelium as a decreasing concentration gradient. Using pharmacological and genetic inhibition, we show that this gradient is created by dual oxidase (Duox), and that it is required for rapid recruitment of leukocytes to the wound. This is the first observation, to our knowledge, of a tissue-scale H_2O_2 pattern, and the first evidence that H_2O_2 signals to leukocytes in tissues, in addition to its known antiseptic role.

Hydrogen peroxide (H_2O_2) is a chemically relatively stable reactive oxygen species (ROS) that can diffuse in tissues and cross cell membranes[2], making it an interesting candidate for paracrine tissue signalling. Plants exploit H_2O_2 as a paracrine signal to regulate xylem differentiation and lignification[2]. The known signalling roles of H_2O_2 in animals are primarily within the cytoplasm, where it regulates metabolism, phosphatase activity and gene transcription, and causes oxidative damage at higher concentrations[3]. Paracrine signalling has been seen in cell culture experiments, but these may not faithfully mimic extracellular conditions in tissues[2].

To investigate possible paracrine signalling by H_2O_2, we imaged its spatiotemporal dynamics, together with leukocyte motility, in an intact vertebrate tissue subjected to mechanical wounding. The zebrafish larval tail fin has become a popular vertebrate model system to study inflammatory and regenerative responses to wounds[4–7]. Rapid leukocyte recruitment to the wound can be easily imaged, and the molecular dynamics of the tissue perturbed using morpholino knockdown, transgenic expression and pharmacology.

We measured H_2O_2 by expressing HyPer, a genetically encoded ratiometric sensor that is highly selective for H_2O_2 over other ROS[1]. HyPer consists of the bacterial H_2O_2-sensitive transcription factor OxyR fused to a circularly permuted yellow fluorescent protein (YFP). Cysteine oxidation of the OxyR part induces a conformational change that increases emission excited at 500 nm (YFP_{500}) and decreases emission excited at 420 nm (YFP_{420}). This change is rapidly reversible within the reducing cytoplasmic environment, allowing dynamic monitoring of intracellular H_2O_2 concentration. We introduced HyPer by messenger RNA injection into zebrafish embryos to induce global cytoplasmic expression (Fig. 1a) and confirmed that HyPer ratios respond to externally added H_2O_2 (Supplementary Fig. 1a).

Upon local injury of the tail fin of zebrafish larvae at 3 days post fertilization (3 d.p.f.), we observed a rapid and marked increase in HyPer ratio signal (YFP_{500}/YFP_{420}) at the wound margin (Fig. 1b and Supplementary Movie 1). To test if this was caused by a non-specific environmental effect (for example, pH change), we expressed YFP alone, and observed only a marginal fluorescence increase, most probably due to ruffling/contraction of the wound margin inducing a local increase in tissue thickness (Supplementary Fig. 1b). Similarly, the pH reporter BCECF-AM did not indicate a major contribution of pH to the wound margin signal. H_2O_2 production at the wound margin was confirmed by using the H_2O_2-selective fluorogenic probe acetyl-pentafluorobenzene sulphonyl fluorescein[8] (Supplementary Fig. 1c). Hence, the primary wound margin signal is due to H_2O_2 or a closely related molecule—HyPer does not respond to superoxide (O_2^-) or nitric oxide (NO)[1]. The H_2O_2 signal peaked ~20 min after wounding (Fig. 1c). At this time, the observable H_2O_2 gradient extended ~100–200 μm inward from the wound margin (Fig. 1d), so that its low concentration end approached the nearest blood vessel.

We quantified leukocyte recruitment to the wound by imaging transmitted light and two different leukocyte-specific fluorescent tags: *mpo::GFP* (ref. 9) (Figs 2d and 3c) and *lysC::DsRED2* (ref. 10) (Fig. 1e). Some leukocytes were located in the fin at the time of wounding, whereas others were apparently recruited from the vasculature. Excluding occasional cases where a leukocyte was already present at the wound site, the first leukocyte arrived at the wound margin 17 ± 6 min after wounding (mean ± s.d. of $n = 14$ larvae). This timing is superimposed on a typical H_2O_2 profile in Fig. 1c. Wound margin H_2O_2 production clearly preceded recruitment of the first leukocyte in most cases (see also Fig. 1e, Supplementary Fig. 1d and Supplementary Movie 2), indicating that the source of H_2O_2 must be tail-fin epithelial cells, not leukocytes. This finding runs counter to the prevailing view that ROS production during inflammatory responses originates from leukocyte oxidative bursts[11].

The main physiological source of extracellular H_2O_2 is likely to be NADPH oxidases, which transport electrons from cytoplasmic NADPH to generate O_2^- or H_2O_2 in phagosomes or outside the cell[12]. The zebrafish genome encodes Nox1, Nox2 (leukocyte oxidase), Nox4, Nox5 and a single isoform of Duox[13] (Fig. 2a). Nox1–5 generate superoxide, which can be dismutated into H_2O_2 by separate superoxide

[1]Department of Systems Biology, Harvard Medical School, Boston, [2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Division of Hematology/Oncology, Department of Pediatrics, Children's Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. †Present address: Karlsruhe Institute of Technology, Forschungszentrum Karlsruhe GmbH, Institute of Toxicology and Genetics, 76344 Eggenstein-Leopoldshafen, Germany.
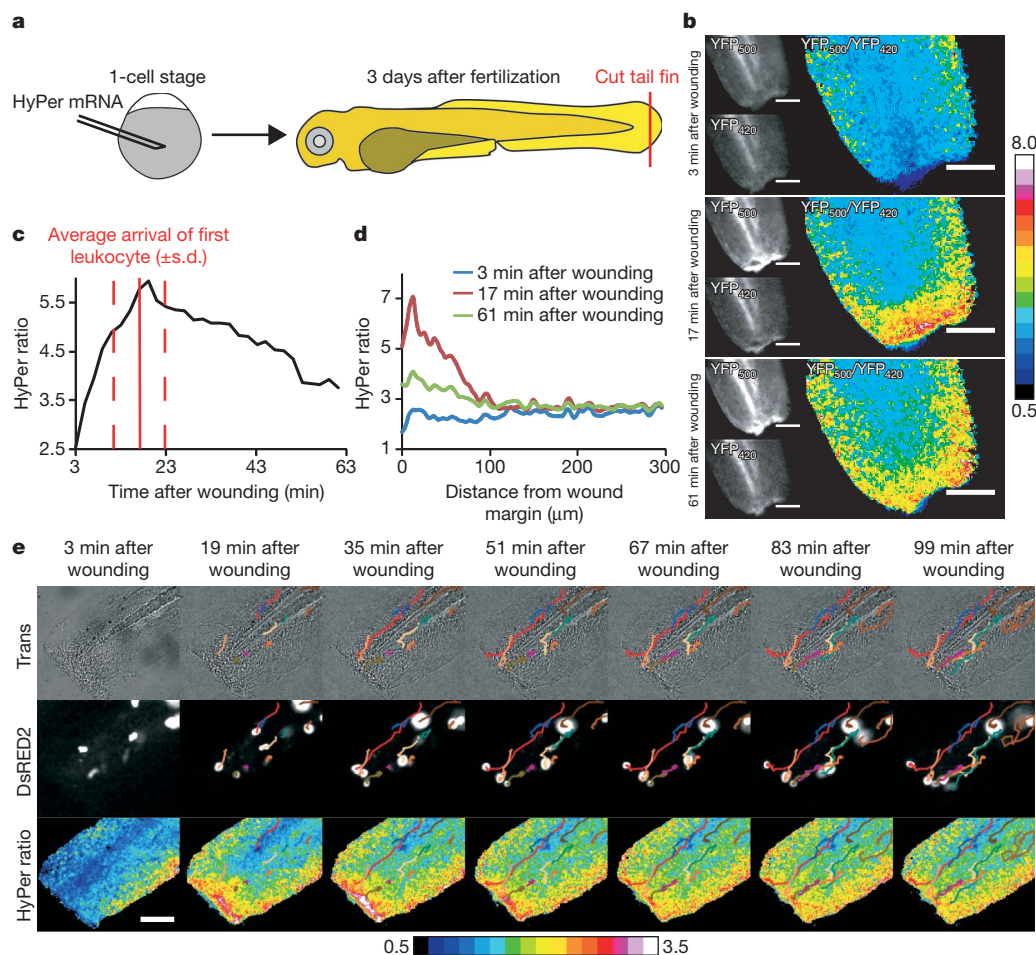*These authors contributed equally to this work.

Figure 1 | Wound margin H_2O_2 production in zebrafish larvae.
a, Experimental procedure. **b**, HyPer imaging in an injured zebrafish larva. $[H_2O_2]$ is inferred from the YFP_{500}/YFP_{420} excitation ratio of HyPer. Greyscale scaling is adjusted to improve contrast. **c**, Temporal $[H_2O_2]$ profile in a ~10–30-μm broad region of interest along the wound margin. Arrival of first leukocyte at wound (solid red line) ± s.d. (dashed red line) is shown. **d**, $[H_2O_2]$ line profile normal to the wound margin. **e**, Imaging of leukocyte recruitment and $[H_2O_2]$ in a *lysC::DsRED2* fish line[10]. Coloured lines are superimposed leukocyte tracks. Pseudo-colour calibration bars: HyPer ratio (YFP_{500}/YFP_{420}). Scale bars: 100 μm.

dismutase (SOD) enzymes, whereas Duox generates H_2O_2 without requiring a separate SOD[14]. To test for a role of any Nox enzyme in generating wound margin H_2O_2, we added two structurally unrelated small molecule inhibitors of the whole family, diphenyleneiodonium (DPI) and VAS2870[15–18], to the bathing water before wounding. Both efficiently inhibited H_2O_2 production without obvious toxicity (Fig. 2b, c; see also Supplementary Movie 3).

We next quantified leukocyte recruitment in the *mpo::GFP* fish line[9] during the peak period of H_2O_2 production. Under control conditions, an average of 4–6 leukocytes arrived at the wound margins within the first 42 min after wounding. Nox inhibition strongly attenuated leukocyte recruitment to the wound during this initial phase of the response, with less than one leukocyte arriving, on average, in drug-treated larvae (Fig. 2d, e and Supplementary Movie 4).

The specific Nox that generates wound margin H_2O_2 was identified by targeting pre-mRNA splice sites with antisense morpholinos. Interference with pre-mRNA splicing of P22phox (*cyba*), an essential subunit of Nox1–4 (ref. 12), led to quantitative conversion of its mRNA level into a mutant with a premature stop codon that probably terminated translation of P22phox after the 28th amino acid residue (Supplementary Fig. 2c, inset). This had no effect on the H_2O_2 gradient (Supplementary Fig. 2a–c and Supplementary Movie 5) or leukocyte recruitment to the wound (Supplementary Fig. 2d). Nox5 and Duox remained as candidates. By semi-quantitative polymerase chain reaction (PCR) we confirmed that *duox* but not *nox5* is expressed in tail-fin tip tissue (Supplementary Fig. 3c). In mammals, DUOX is

mainly expressed in the thyroid gland, where it generates H_2O_2 for organification of I^- (ref. 19), but also in epithelial surfaces that contact liquid environments, including the luminal surface of the gut and lung. Extracellular H_2O_2 made by DUOX is thought to react with halide or thiocyanate, catalysed by secreted lactoperoxidases, to generate more ROS that kill luminal bacteria[20,21]. DUOX contains two Ca^{2+} binding EF-hand motifs and, at least in cell culture, can be activated by Ca^{2+}-mobilizing small molecules[22] and by mechanical cell injury[23], making it a good candidate for wound signalling.

Morpholino-induced perturbation of *duox* pre-mRNA splicing (Fig. 3b, inset) caused a developmental morphology phenotype characterized by cell death predominantly in the head region. This phenotype is probably specific for *duox* knockdown, because two independent splice morpholinos, but not a corresponding 5-misprime morpholino, induced the same phenotype (not shown). To generate morphologically normal tail fins for the assay, we simultaneously knocked down *p53*, which partially rescued the *duox* knockdown morphological phenotype. Notably, we found a significant reduction of wound-induced H_2O_2 production in *duox/p53* knockdown larvae compared with *duox 5-MP/p53* controls (Fig. 3a, b, Supplementary Fig. 2a and Supplementary Movie 6). Furthermore, *duox* knockdown strongly attenuated recruitment of leukocytes to the wound (Fig. 3c, d and Supplementary Movies 7 and 8). This attenuation was not caused by a reduction of total leukocyte number in *duox* knockdown larvae (Supplementary Fig. 2e). *Duox* knockdown did cause a significant reduction in the number of
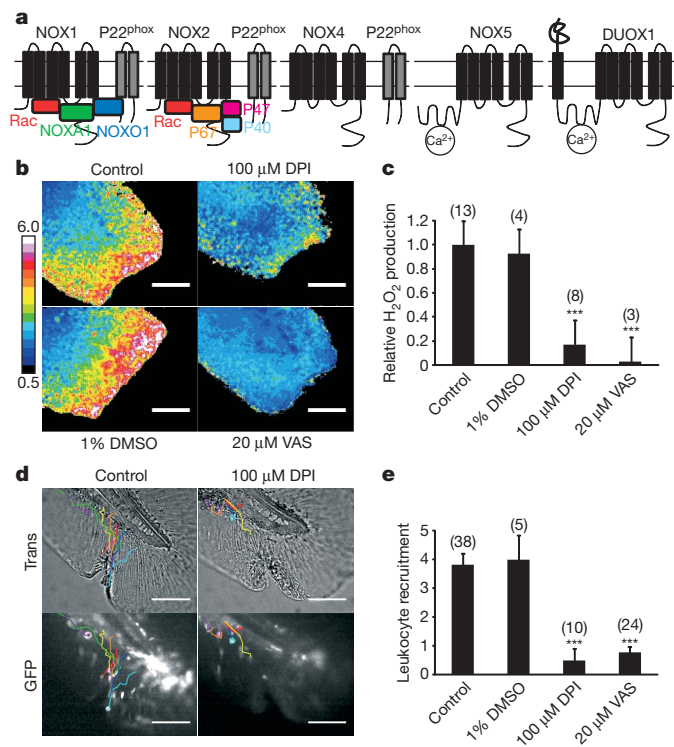
Figure 2 | Nox/Duox activity is required for wound margin H$_2$O$_2$ production and leukocyte recruitment. a, Scheme of mammalian NADPH oxidases also found in zebrafish[12,13]. **b**, Wound margin [H$_2$O$_2$] with/without DPI or VAS2870 (VAS), or carrier (1% DMSO), imaged 17 min after wounding. Pseudo-colour calibration bar: HyPer ratio (YFP$_{500}$/YFP$_{420}$). **c**, Statistical quantification of wound margin [H$_2$O$_2$]. **d**, Injured tail fins of *mpo::GFP* larvae[9] with/without DPI (42 min after wounding). Coloured lines are leukocyte tracks derived from the corresponding time-lapse movies. **e**, Statistical quantification of leukocyte recruitment to wound margin. Error bars indicate s.e.m. of indicated number of larvae (brackets). Triple asterisk, $P < 0.001$ (versus control). Scale bars: 100 μm.

leukocytes infiltrating the tail fin after wounding, reducing it to near the level seen in unwounded fins (Supplementary Fig. 3a). It also caused a significant decrease in directional migration towards the wound, whereas basal leukocyte motility, as observed in the absence of a wound, was not affected. These data implicate Duox as the main source of wound margin H$_2$O$_2$ required for rapid leukocyte recruitment.

We have visualized, for the first time, a tissue-scale gradient of H$_2$O$_2$ induced by wounding, found that it is generated by Duox activity in epithelial cells, and shown that it is required for leukocyte recruitment to the wound. On the basis of published calibration of HyPer in tissue culture[1], wound-induced extracellular H$_2$O$_2$ may reach concentrations of ~0.5–50 μM near the wound margin. The gradient was established within 10 min of wounding, and gradually dissipated over ~1–2 h. Visual inspection of movies (Figs 1e, 2d and 3c) suggested that leukocytes sensed the wound within ~10 min, from distances as large as 200 μm. Thus, the spatiotemporal scales of the H$_2$O$_2$ gradient and the leukocyte response were roughly similar. Trajectory analysis showed that the H$_2$O$_2$ gradient stimulated leukocyte recruitment mainly by increasing directionality of leukocyte migration and tissue infiltration (Supplementary Fig. 3a and Supplementary Movies 4 and 8). This argues against a permissive role of extracellular H$_2$O$_2$ for basal leukocyte motility in our assay, and favours the idea that wound margin H$_2$O$_2$ production spatially instructs rapid wound recruitment of leukocytes, either by direct chemotactic signalling or by stimulating production of a downstream chemoattractant. Direct chemotactic activity of H$_2$O$_2$ was previously demonstrated *in vitro* for neutrophils[24] and vascular smooth muscle
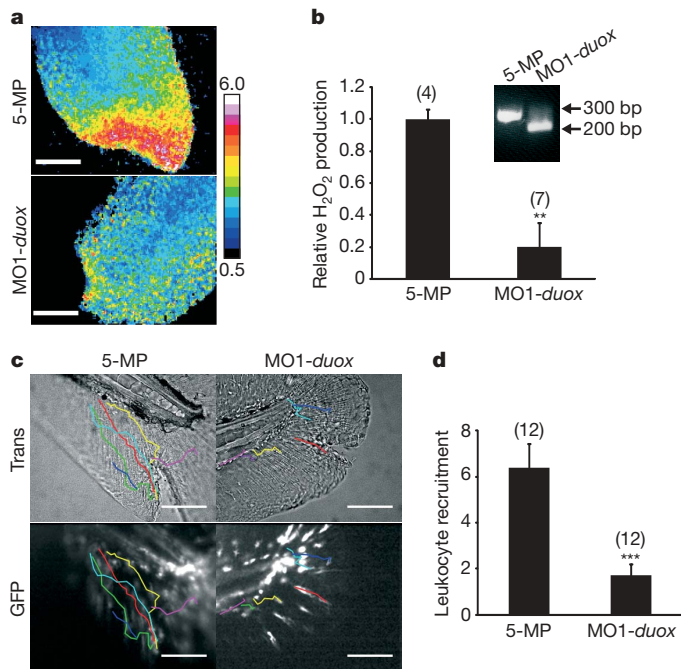

Figure 3 | Duox activity is required for wound margin H$_2$O$_2$ production and leukocyte recruitment. a, Wound margin H$_2$O$_2$ after morpholino-mediated *duox* knockdown (MO1-*duox*) or injection of a corresponding 5-misprime morpholino (5-MP) imaged 17 min after wounding. Inset: RT–PCR of a *duox* mRNA region flanking the targeted splice site. Pseudo-colour calibration bar: HyPer ratio (YFP$_{500}$/YFP$_{420}$). **b**, Quantification of wound margin [H$_2$O$_2$]. **c**, Injured tail fins of *mpo::GFP* larvae[9] injected with MO1-*duox*, or 5-MP (42 min after wounding). Coloured lines are leukocyte tracks. **d**, Quantification of leukocyte recruitment. Error bars indicate s.e.m. of indicated number of larvae (brackets). Double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ (versus control). Scale bars: 100 μm.

cells[25] at concentrations[24] (~10 μM) that are roughly consistent with our estimation of wound margin [H$_2$O$_2$]. Together with our data, this raises the possibility that H$_2$O$_2$ itself acts as a paracrine, chemotactic signal during the initial phase of wound detection. Leukocytes might express trans-membrane receptors for H$_2$O$_2$—none are known, but T-type Ca^{2+} channels are thought to have this function in sensory neurons[26]. Alternatively, H$_2$O$_2$ might direct migration by entering the cytoplasm and locally modifying intracellular receptors, such as the redox-sensitive phosphatase PTEN[27]. Phosphatidylinositol-3,4,5-trisphosphate phosphatases such as PTEN or SHIP1 are thought to be important regulators of chemo- and electrotactic responses[28–30]. Our current data do not distinguish whether the spatial H$_2$O$_2$ gradient reflects diffusion from a localized source at the wound margin combined with global breakdown by catalases and/or peroxidases, or rather a gradient of H$_2$O$_2$ production induced by some upstream regulatory pattern, such as an electric field or a spatial gradient of an upstream signalling molecule. DUOX was previously implicated in constitutive ROS-induced microbial killing by mucosal epithelia[19]. Our data implicate it, for the first time, as a major, non-myeloid ROS source in the initial phase of inflammation. We hypothesize that the DUOX/lactoperoxidase system evolved to have two useful roles in early responses to epithelial wounding: local killing of invading bacteria, and rapid recruitment of phagocytic leukocytes from distant sites.

METHODS SUMMARY

Imaging of H$_2$O$_2$ and leukocytes in zebrafish. One-cell-stage zebrafish embryos were injected with HyPer mRNA. 3 d.p.f. larvae were subjected to tail-fin tip amputation and mounted in 1% low-melting agarose. HyPer fluorescence was excited with 501/16 and 420/40 band-pass excitation filters and corresponding YFP emission was acquired every 2 min within 3–42 min after injury using a 535/30 band-pass emission filter. For calculating HyPer ratio images, YFP$_{500}$ and

YFP_{420} images were smoothed, background subtracted, thresholded, and then divided.

Leukocytes were imaged every 30 s within 3–42 min after tail-fin incision of fluorescent leukocyte reporter zebrafish larvae. Leukocyte migration to the wound was observed both by fluorescence and transmission imaging. Final leukocyte count at the wound margin was assessed 42 min after injury.

Zebrafish larvae were anaesthetized for wounding and imaging experiments. All media were sterile filtered. Blades were treated with 70% ethanol before use. Imaging was optimized for low illumination, and performed on an inverted wide-field microscope equipped with a CCD camera and a mercury illumination source.

Genetic and pharmacological perturbations. Anaesthetized larvae were incubated with pharmacological compounds up to 40 min before wounding and during imaging. Antisense morpholinos were injected into 1-cell-stage embryos. Morpholino-mediated splice perturbation was confirmed by RT–PCR.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Belousov, V. V. et al. Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. Nature Methods 3, 281–286 (2006).
2. Bienert, G. P., Schjoerring, J. K. & Jahn, T. P. Membrane transport of hydrogen peroxide. Biochim. Biophys. Acta 1758, 994–1003 (2006).
3. Oktyabrsky, O. N. & Smirnova, G. V. Redox regulation of cellular functions. Biochemistry 72, 132–145 (2007).
4. Redd, M. J., Cooper, L., Wood, W., Stramer, B. & Martin, P. Wound healing and inflammation: embryos reveal the way to perfect repair. Phil. Trans. R. Soc. Lond. B 359, 777–784 (2004).
5. Renshaw, S. A., Loynes, C. A., Elworthy, S., Ingham, P. W. & Whyte, M. K. Modeling inflammation in the zebrafish: how a fish can help us understand lung disease. Exp. Lung Res. 33, 549–554 (2007).
6. Grabher, C. et al. Birth and life of tissue macrophages and their migration in embryogenesis and inflammation in medaka. J. Leukoc. Biol. 81, 263–271 (2007).
7. Huttenlocher, A. & Poznansky, M. C. Reverse leukocyte migration can be attractive or repulsive. Trends Cell Biol. 18, 298–306 (2008).
8. Maeda, H. et al. Fluorescent probes for hydrogen peroxide based on a non-oxidative mechanism. Angew. Chem. Int. Edn Engl. 43, 2389–2391 (2004).
9. Mathias, J. R. et al. Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish. J. Leukoc. Biol. 80, 1281–1288 (2006).
10. Hall, C., Flores, M. V., Storm, T., Crosier, K. & Crosier, P. The zebrafish lysozyme C promoter drives myeloid-specific expression in transgenic fish. BMC Dev. Biol. 7, 42 (2007).
11. Sen, C. K. & Roy, S. Redox signals in wound healing. Biochim. Biophys. Acta 1780, 1348–1361 (2008).
12. Bedard, K. & Krause, K. H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. Physiol. Rev. 87, 245–313 (2007).
13. Kawahara, T., Quinn, M. T. & Lambeth, J. D. Molecular evolution of the reactive oxygen-generating NADPH oxidase (Nox/Duox) family of enzymes. BMC Evol. Biol. 7, 109 (2007).
14. Ameziane-El-Hassani, R. et al. Dual oxidase-2 has an intrinsic Ca^{2+}-dependent H_2O_2-generating activity. J. Biol. Chem. 280, 30046–30054 (2005).
15. ten Freyhaus, H. et al. Novel Nox inhibitor VAS2870 attenuates PDGF-dependent smooth muscle cell chemotaxis, but not proliferation. Cardiovasc. Res. 71, 331–341 (2006).
16. Stielow, C. et al. Novel Nox inhibitor of oxLDL-induced reactive oxygen species formation in human endothelial cells. Biochem. Biophys. Res. Commun. 344, 200–205 (2006).
17. Lange, S. et al. Platelet-derived growth factor BB stimulates vasculogenesis of embryonic stem-cell-derived endothelial cells by calcium-mediated generation of reactive oxygen species. Cardiovasc. Res. 81, 159–168 (2009).
18. Tegtmeier, F. et al. Compounds containing a N-heteroaryl moiety linked to fused ring moieties for the inhibition of NAD(P)H oxidases and platelet activation. Patent WO/2005/111041 (2005).
19. Donko, A., Peterfi, Z., Sum, A., Leto, T. & Geiszt, M. Dual oxidases. Phil. Trans. R. Soc. Lond. B 360, 2301–2308 (2005).
20. Geiszt, M., Witta, J., Baffi, J., Lekstrom, K. & Leto, T. L. Dual oxidases represent novel hydrogen peroxide sources supporting mucosal surface host defense. FASEB J. 17, 1502–1504 (2003).
21. Ha, E. M., Oh, C. T., Bae, Y. S. & Lee, W. J. A direct role for dual oxidase in Drosophila gut immunity. Science 310, 847–850 (2005).
22. Forteza, R., Salathe, M., Miot, F., Forteza, R. & Conner, G. E. Regulated hydrogen peroxide production by Duox in human airway epithelial cells. Am. J. Respir. Cell Mol. Biol. 32, 462–469 (2005).
23. Wesley, U. V., Bove, P. F., Hristova, M., McCarthy, S. & van der Vliet, A. Airway epithelial cell migration and wound repair by ATP-mediated activation of dual oxidase 1. J. Biol. Chem. 282, 3213–3220 (2007).
24. Klyubin, I. V., Kirpichnikova, K. M. & Gamaley, I. A. Hydrogen peroxide-induced chemotaxis of mouse peritoneal neutrophils. Eur. J. Cell Biol. 70, 347–351 (1996).
25. Li, W., Liu, G., Chou, I. N. & Kagan, H. M. Hydrogen peroxide-mediated, lysyl oxidase-dependent chemotaxis of vascular smooth muscle cells. J. Cell. Biochem. 78, 550–557 (2000).
26. Todorovic, S. M. et al. Redox modulation of T-type calcium channels in rat peripheral nociceptors. Neuron 31, 75–85 (2001).
27. Kwon, J. et al. Reversible oxidation and inactivation of the tumor suppressor PTEN in cells stimulated with peptide growth factors. Proc. Natl Acad. Sci. USA 101, 16419–16424 (2004).
28. Zhao, M. et al. Electrical signals control wound healing through phosphatidylinositol-3-OH kinase-γ and PTEN. Nature 442, 457–460 (2006).
29. Subramanian, K. K. et al. Tumor suppressor PTEN is a physiologic suppressor of chemoattractant-mediated neutrophil functions. Blood 109, 4028–4037 (2007).
30. Nishio, M. et al. Control of cell polarity and motility by the $PtdIns(3,4,5)P_3$ phosphatase SHIP1. Nature Cell Biol. 9, 36–44 (2007).

Author Contributions P.N. and T.J.M. conceived the project. P.N. and C.G. designed and executed the experiments. C.G. and A.T.L. contributed expertise in the zebrafish system. P.N. and T.J.M. contributed expertise in imaging and pharmacology. T.J.M. and A.T.L. provided guidance and institutional support. P.N., C.G. and T.J.M. wrote the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.N. (Philipp_Niethammer@hms.harvard.edu).

METHODS

General fish procedures. Zebrafish strains AB, *mpo::GFP* (ref. 9) and *lysC::DsRED2* (ref. 10) were maintained as described[31]. For wounding assays, zebrafish were anaesthetized in E3 (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl$_2$, 0.33 mM MgSO$_4$) containing 0.1 mg ml^{-1} Tricaine (Sigma) before wounding. To prevent pigment formation, larvae were maintained in E3 containing 0.2 mM *N*-phenylthiourea (PTU; Sigma). All buffers were sterile filtered and blades were sterilized using 70% ethanol before use.

Imaging the wound response. For all imaging, larvae were maintained in E3 supplemented with 0.1 mg ml^{-1} Tricaine.

For imaging of H$_2$O$_2$ production, 1-cell-stage zebrafish embryos were injected with HyPer[1] mRNA (~0.5 mg ml^{-1}). 3 d.p.f. larvae were subjected to tail-fin tip amputation using a needle knife (Fine Science Tools) and embedded in 1% low melting agarose (Lonza) in a glass-bottom dish (Matek Corporation). Every 2 min starting ~3 min after wounding, HyPer fluorescence was excited using 420/40 and 501/16 band-pass filters (Chroma), and YFP emission was acquired using a 535/30 band-pass filter.

For alternative H$_2$O$_2$ imaging, 2–3 d.p.f. larvae were loaded ~60 min with 50 μM acetyl-pentafluorobenzene sulphonyl fluorescein (Calbiochem) before wounding. Emission was excited using a 484/15 band-pass filter (Chroma) and acquired using a 525/50 band-pass filter (Chroma).

For pH imaging, 2–3 d.p.f. larvae were loaded ~60 min with 50 μM 3′-*O*-acetyl-2′,7′-bis(carboxyethyl)-4 or 5-carboxyfluorescein diacetoxymethyl ester (BCECF-AM, Calbiochem) before wounding. Emission was excited using a 484/15 band-pass filter (Chroma) and acquired using a 535/30 band-pass filter (Chroma). For imaging of leukocyte recruitment, 3–5 d.p.f. *mpo::GFP* or *lysC::DsRED2* larvae were cut at the tail fin using a tungsten needle (Fine Science Tools), mounted in agarose, and leukocyte fluorescence was excited using a 484/15 or 540/15 band-pass filter (Chroma). Emission was acquired every 30 s using a 525/50 or 610/80 band-pass filter (Chroma). All images were acquired at room temperature (~26 °C) using Metamorph (Molecular Devices) and a Nikon Eclipse TE300 microscope equipped with 20× plan-apochromate NA 0.75 air objective lens, an ORCA-ER camera (Hamamatsu), and a mercury light source (Chiu Technical Corporation).

Tail-fin tip amputation (needle blade) was used in all HyPer assays; tail-fin incision (tungsten needle) in all chemotaxis assays. Only the same types of cuts were directly compared. The HyPer signal was not dependent on the type of cut (Supplementary Fig. 3b).

Generally, each imaging set-up was optimized for minimal light exposure of larvae. Whenever possible, we used two neutral density filters (except for BCECF-AM where dye loading was rather inefficient, so that we had to use one neutral density filter), highest camera gain, and high binning (for example, 8× bin for probe imaging, 4× bin for leukocyte imaging).

Image processing and data analysis. For calculating HyPer ratio images, smoothed (one-pass median filter), background subtracted and thresholded YFP$_{500}$ and YFP$_{420}$ images were divided (YFP$_{500}$/YFP$_{420}$) using ImageJ (NIH) or Matlab (Mathworks). Upregulation of H$_2$O$_2$ was calculated by dividing the mean ratio acquired in a region of interest directly at the wound margin (r_{wound}) by the mean basal ratio acquired in a region of interest inside the body (r_{basal}, ~300–400 μm distant from the wound margin). H$_2$O$_2$ upregulation was expressed either as multiple ($f_{mult} = r_{wound}/r_{basal}$; for example, Supplementary Fig. 2b) or fraction of the base level ($f_{fract} = (r_{wound} - r_{basal})/r_{basal}$). Relative H$_2O_2$ production at wound margin (for example, Figs 2c and 3b) was derived as f_{fract} normalized to the control ($f_{rel} = f_{fract}(\text{sample})/f_{fract}(\text{control})$).

Leukocyte recruitment was determined by counting all migrating cells that arrived at the wound margin within 42 min after wounding as judged from the 30 s per frame time-lapse movies of *mpo::GFP* leukocyte reporter fish (transmission and GFP channel). Cells that already resided at the wound margin at the beginning of the time-lapse sequence (~3 min after wounding) were not counted.

Leukocyte trajectory analysis. Trajectory analysis was performed on the same leukocyte time-lapse data that were also used to quantify wound recruitment of leukocytes. Trajectories were generated by marking the approximate centre of mass of those cells that moved in the ventral tail fin, and could be identified with adequate reliability. Only those cells were included in the statistical path analysis that described a path of at least 50 μm. Furthermore, tracks or part of tracks

within a radius of 50 μm around the centre of mass of the triangular wound region were not included into the analysis to avoid tracking of cells that had already reached the wound, and merely moved along the wound margins. Average velocity (v_{av}) was calculated as $v_{av} = l_{track}/t_{track}$, with l_{track} being the length of the track and t_{track} being the total track time.

Path linearity (which is frequently also termed 'directionality') was calculated as $\text{Dir}_p = d_{OE}/l_{track}$, with d_{OE} being the Euclidian distance between origin (O) and endpoint (E) of the track.

Wound directionality was calculated as $\text{Dir}_w = (d_{OW} - d_{EW})/l_{track}$, with d_{OW} being the distance between track origin and centre of mass of the wound (W), and d_{EW} being the distance between track endpoint and W.

Pharmacological and morpholino treatments. Larvae were incubated in E3 supplemented with 100 μM DPI (Sigma), 20 μM VAS2870, 1% DMSO (Sigma) 30–40 min before wounding. Mounting agarose and imaging medium (E3) were supplemented with the indicated compound concentrations.

The following splice morpholinos (Gene Tools) were injected into 1-cell-stage larvae (~0.5–1 mM): MO-*cyba*, 5′- ATCATAGCATGTAAGGATACATCCC-3′; MO1-*duox*, 5′-AGTGAATTAGAGAAATGCACCTTTT-3′; MO5-MP-*duox* (5-MP), 5′- AGTcAATTAcAGAAATcCAgCTaTT-3′; MO2-*duox*, 5′- ACATTC ACTCTCTCACCTGGATATG-3′. Lower-case letters indicate base-pair mismatch with the target sequence.

For morphotyping, RNA was prepared from 3 d.p.f. larvae by phenol-chloroform extraction (TRI solution, Ambion), and one-step RT–PCR (Qiagen) was performed to confirm knockdown efficiency using the following primers: *cyba* forward, 5′-GCGAAGATTGAGTGGGCGATGTGGGCC-3′; *cyba* reverse, 5′-TTATTCGTTGATGGTGACAGACATAGGATTGTC-3′; *duox* forward, 5′-ACACATGTGACTTCATATCCAG-3′; *duox* reverse, 5′-ATTATTAAC TCATCCACATCCAG-3′.

The RT–PCR products were sequenced. MO-*cyba*-mediated splice perturbation produced a 146-bp deletion in the *cyba* mRNA, introducing a premature stop codon into the resulting splice-mutant mRNA, coding for a 28-amino-acid truncated translation product. MO1-*duox*-mediated splice perturbation produced a 39-bp in-frame deletion within the peroxidase-like domain of *duox*. MO1-*duox* and MO2-*duox* produced identical phenotypes; however, MO2-*duox*-injected larvae yielded neither detectable wild-type nor splice-mutant amplification product, whereas *beta-actin* could successfully be amplified from the same template. This indicated that MO2-*duox* resulted in *duox* mRNA knockdown, either by generating a splice-mutant mRNA that was rapidly degraded or that was too large to be amplified under our conditions. For phenotypical rescue of the tail fin in the HyPer and leukocyte migration assays, MO1-*duox* and MO5-MP-*duox* (5-MP) morpholinos were generally co-injected with a morpholino inhibiting *p53* mRNA translation (~0.2 mM, 5′-GCG CCATTGCTTTGCAAGAATTG-3′[32]).

Cell sorting. Larvae (~150) were collected from Tg(*mpo::GFP*) at 80 h.p.f. and disaggregated into a single-cell suspension as previously described[33]. Sorting of *mpo::GFP* positive cells was performed on a BD Aria based on GFP fluorescence.

RNA isolation and semi-quantitative RT–PCR. Total RNA from amputated tail fins (80 h.p.f.) or sorted GFP$^+$ cells was extracted with TRIzol (Invitrogen). RT–PCR was performed with a one-Step RT–PCR kit (Qiagen) according to the manufacturer's protocol using 35 cycles on 2 ng total RNA with intron-spanning primers. Oligo sequences were as follows: *duox* forward, 5′-GTTGGCTT TGGTGTAACTGTA-3′; *duox* reverse, 5′-GCCCAGGCTGTGAGAG-3′; *nox5* forward, 5′-TGGCCTAATGGTGGTCTGTTC-3′; *nox5* reverse, 5′-CAGAGC CGAAACCCAGATG-3′; *beta-actin* forward, 5′-CATTGGCAATGAGCGTT TC-3′; *beta-actin* reverse, 5′-TACTCCTGCTTGCTGATCCAC-3′.

Statistics. All error bars indicate standard errors of means (s.e.m.). All *P*-values have been derived by an unpaired, two-tailed *t*-test assuming unequal variances (heteroscedastic) using Excel (Microsoft).

31. Nusslein-Volhard, C. & Dahm, R. Zebrafish (Oxford Univ. Press, 2002).
32. Chen, J. *et al.* Loss of function of def selectively up-regulates Delta113p53 expression to arrest expansion growth of digestive organs in zebrafish. *Genes Dev.* **19**, 2900–2911 (2005).
33. Bertrand, J. Y. *et al.* Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. *Development* **134**, 4147–4156 (2007).

LETTERS

CCR7 signalling as an essential regulator of CNS infiltration in T-cell leukaemia

Silvia Buonamici[1,2], Thomas Trimarchi[1,2], Maria Grazia Ruocco[1,3], Linsey Reavie[1,2], Severine Cathelin[1,2], Brenton G. Mar[4], Apostolos Klinakis[5], Yevgeniy Lukyanov[1], Jen-Chieh Tseng[1], Filiz Sen[1,2], Eric Gehrie[6], Mengling Li[7], Elizabeth Newcomb[1], Jiri Zavadil[1], Daniel Meruelo[1], Martin Lipp[8], Sherif Ibrahim[1], Argiris Efstratiadis[5], David Zagzag[1], Jonathan S. Bromberg[6], Michael L. Dustin[1,3] & Iannis Aifantis[1,2]

T-cell acute lymphoblastic leukaemia (T-ALL) is a blood malignancy afflicting mainly children and adolescents[1]. T-ALL patients present at diagnosis with increased white cell counts and hepatosplenomegaly, and are at an increased risk of central nervous system (CNS) relapse[2,3]. For that reason, T-ALL patients usually receive cranial irradiation in addition to intensified intrathecal chemotherapy. The marked increase in survival is thought to be worth the considerable side-effects associated with this therapy. Such complications include secondary tumours, neurocognitive deficits, endocrine disorders and growth impairment[3]. Little is known about the mechanism of leukaemic cell infiltration of the CNS, despite its clinical importance[4]. Here we show, using T-ALL animal modelling and gene-expression profiling, that the chemokine receptor CCR7 (ref. 5) is the essential adhesion signal required for the targeting of leukaemic T-cells into the CNS. *Ccr7* gene expression is controlled by the activity of the T-ALL oncogene Notch1 and is expressed in human tumours carrying Notch1-activating mutations. Silencing of either CCR7 or its chemokine ligand CCL19 (ref. 6) in an animal model of T-ALL specifically inhibits CNS infiltration. Furthermore, murine CNS-targeting by human T-ALL cells depends on their ability to express CCR7. These studies identify a single chemokine–receptor interaction as a CNS 'entry' signal, and open the way for future pharmacological targeting. Targeted inhibition of CNS involvement in T-ALL could potentially decrease the intensity of CNS-targeted therapy, thus reducing its associated short- and long-term complications.

Recent studies have shown that mutations of the developmental regulator Notch1 can be identified in most T-ALL patients[7]. It is estimated that activation of the Notch1 signalling pathway occurs in at least 80% of all T-ALL cases[7–10]. To investigate the mechanisms of T-ALL CNS infiltration and derive information that could be useful for treatment, we have attempted to establish animal models involving expression of oncogenic Notch1 (intracellular Notch1 fragment, Notch1-IC). The first model entails the transplantation of wild-type haematopoietic progenitors carrying Notch1-IC introduced by retroviral transfer (WT^Notch-IC)[11]. The second model is on the basis of Cre/*loxP* recombination and involves Mx-Cre mice crossed with partners carrying dormant transgenic Notch1-IC, which was knocked-in into the ubiquitously expressed *Eef1a1* locus[12]. The dormant Notch1-IC exerts oncogenic action after excision of a DNA segment blocking its expression, when Cre is expressed in haematopoietic progenitor cells by the IFN-α-inducible Mx1 promoter after polyinosinic:polycytidylic acid (poly(I:C)) injection. Both models developed T-ALL, presented atypical CD4[+] CD8[+] T cells in the peripheral blood as well as characteristic pathological features of T-ALL (Fig. 1 and Supplementary Figs 1 and 2). Immunohistochemical analysis demonstrated that in both models Notch1-IC–EGFP[+] (enhanced green fluorescent protein) and CD3[+] leukaemic cells efficiently infiltrated the leptomeningeal spaces of the brain (Fig. 1b, c and Supplementary Fig. 1). Further studies showed that the CNS infiltration was progressive, and was initially detected in mice in which leukaemic blasts were readily detected in their peripheral blood (Supplementary Fig. 3) and secondary lymphoid tissue (data not shown). We were thus able to show that oncogenic Notch1-IC expression was able to induce T-ALL and target the transformed cells to the CNS.

We used a genome-wide transcriptome approach to identify Notch1-induced adhesion regulators that could be essential for CNS infiltration. Uncommitted haematopoietic progenitors were infected with Notch1-IC–EGFP[+] retroviruses and gene expression was recorded 48 h later[11]. Detailed data mining demonstrated that a considerable fraction of Notch-controlled genes are potential regulators of cell adhesion, migration and metastasis (Fig. 2a and Supplementary Table 1). The expression of a specific gene, the chemokine receptor *Ccr7*, was significantly upregulated (Fig. 2a, b), and its expression remained constant after several days of culture (data not shown). CCR7 overexpression and function was also confirmed by real-time PCR, flow cytometry analysis and *in vitro* chemotaxis assays towards its known chemokine ligands CCL19 and CCL21 (Fig. 2b–d). CCR7 is an attractive candidate because it is a known regulator of lymphocyte migration[6] and has been suggested to be important for the trafficking of lymphocytes participating in CNS immunosurveillance[13,14]. CCR7 functions through its interactions with CCL19 and CCL21 (ref. 6), and the expression and function of all three have been shown to be involved in the directional metastasis of several types of solid tumours, including melanomas and breast cancers[15,16].

To correlate Notch1 activation and CCR7 expression further, we analysed T-ALL lines containing Notch1-activating mutations and primary T-ALL samples. Notably, surface CCR7 was expressed in 80% (4 out of 5) of T-ALL lines and in 73% (8 out of 11) of peripheral blood from primary T-ALL samples (Supplementary Fig. 4). CCR7 expression in the studied lines was dependent on Notch1 activation, as the repression of Notch1 processing due to the addition of γ-secretase
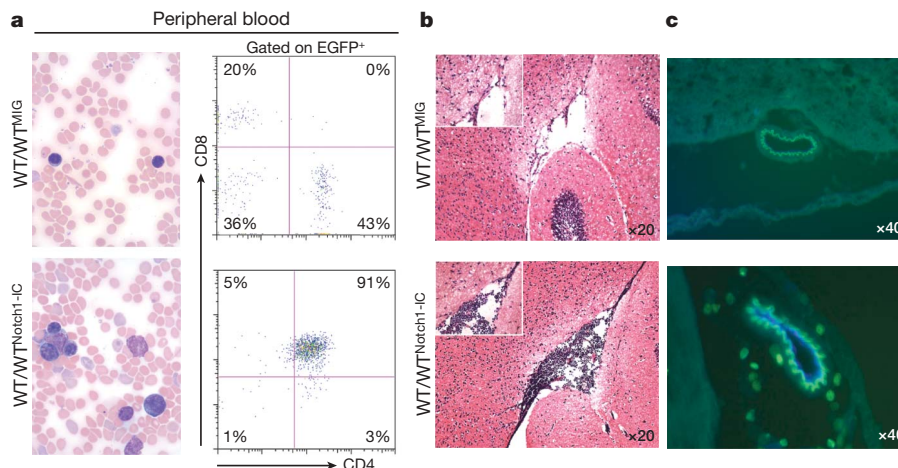
Figure 1 | Notch1 activation induces T-ALL and targets leukaemic cells into the CNS. a, Induction of T-ALL in a transplantation model (WT/ $WT^{Notch1-IC}$). Peripheral blood smears (left), and fluorescence-activated cell sorting (FACS, right) analysis using CD4 and CD8 antibodies are shown. 'WT^{MIG}' denotes wild-type bone marrow infected with a control MIG retrovirus. **b**, Notch1-IC$^+$ EGFP$^+$ cells in the brain meningeal spaces of transplanted mice. **c**, Infiltrating lymphocytes surrounding a brain vessel in leukaemic (bottom panel) but not in healthy (control, top panel) recipients. Co-staining with CD31 antibodies (blue) indicates endothelial cells within the infiltrating lymphocytes.

inhibitors (DBZ or compound E) led to significant downregulation of CCR7 messenger RNA and protein expression (Supplementary Fig. 5 and not shown). To obtain a preliminary estimate of the importance of CCR7 expression in CNS infiltration, we selected two human T-ALL lines (Supplementary Fig. 4a), one that expressed (CEM/ CCR7$^+$) and one that did not express (DND41/CCR7$^-$) the chemokine receptor, and infected them using a luciferase-expressing lentivirus. To study disease induction and progression, we transplanted the cells into alymphoid $Rag2^{-/-}$ $Il2rg^{-/-}$ hosts. We were able to demonstrate that the transplantation of an identical number of leukaemia cells led to distinct survival patterns, as hosts that received CEM/ CCR7$^+$ cells succumbed to the disease earlier than those receiving

the DND41/CCR7$^-$ cells (Fig. 3a). Using live animal bioluminescent imaging, we were able to demonstrate a higher tumour load in mice that received the CEM/CCR7$^+$ cell line (Fig. 3b). Most importantly, the brain and spinal cord of hosts transplanted with CEM/CCR7$^+$ (but not with DND41/CCR7$^-$) were infiltrated by T-ALL cells. Further histopathological analysis confirmed the localization of the infiltrating cells in leptomeningeal spaces (Fig. 3c, d).

As a next step, we addressed the sufficiency of CCR7 overexpression for the recruitment of leukaemic T cells in the CNS. We expressed mouse CCR7 (mCCR7) in the DND41/CCR7$^-$ human T-ALL cell line using a bicistronic retroviral vector expressing CCR7 in conjunction with expression of the mCherry fluorochrome

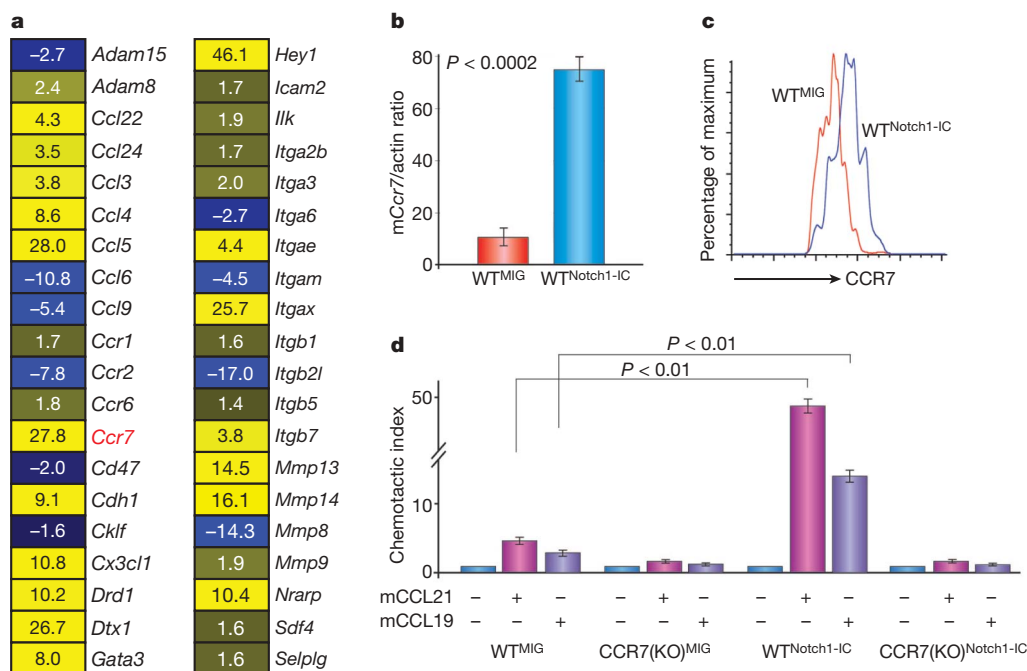**Figure 2 | CCR7 expression and response to CCL19/CCL21 is induced by Notch1 activation. a**, Heat diagram of selected adhesion/migration regulators that are controlled by Notch1-IC. A few classical Notch targets (*Dtx1*, *Gata3*, *Hey1* and *Nrarp*) are also included. For all genes, $P < 0.001$. Yellow and blue denote increased and decreased mRNA abundance, respectively. **b**, **c**, Real-time PCR (**b**) and FACS (**c**) analysis showing the

induction of CCR7 gene and protein expression in haematopoietic progenitors in response to Notch1-IC expression; $n = 4$. **d**, Notch1-IC expression induces the chemotaxis of wild-type (WT$^{Notch1-IC}$), but not CCR7(KO)$^{Notch1-IC}$ progenitors towards both CCL19 and CCL21; $n = 3$. Error bars define s.d. for all experiments.
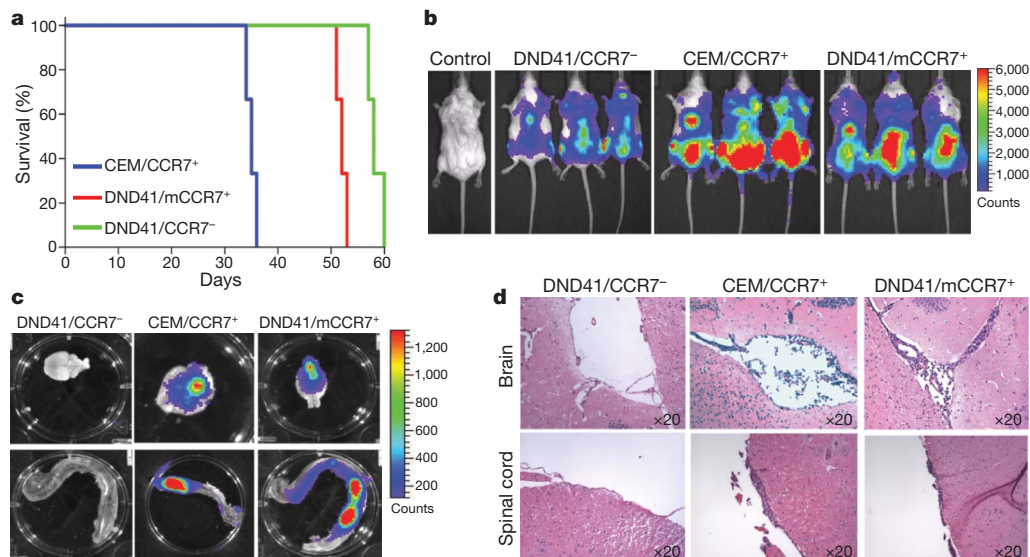
Figure 3 | CCR7 expression is sufficient for CNS infiltration of human T-ALL cells. a, Kaplan–Meyer analysis of recipients that received identical numbers of CEM, DND41 or DND41/mCCR7$^+$ cells; $n = 5$. **b**, Bioluminescent imaging of mice 2 weeks after transplantation with the indicated cell lines. **c**–**d**, Infiltration of T-ALL cells into the CNS of recipient mice as shown using bioluminescence and histochemistry.

(Supplementary Fig. 6a). We verified that retroviral transduction induced mCherry and mCCR7 expression, and that mCherry$^+$ DND41 (DND41/mCCR7$^+$) cells acquired the ability to efficiently respond to both CCL19 and CCL21 ligands (Supplementary Fig. 6b). Infected cells and control (mCherry-only expressing counterparts) were injected into $Rag2^{-/-}$ $Il2rg^{-/-}$ hosts. All recipients developed T-ALL (Fig. 3a, b) and no CNS infiltration was detected in the mice transplanted with the control DND41/CCR7$^-$ cells. On the other hand, we were able to demonstrate that CCR7 ectopic expression was sufficient to allow these cells to infiltrate both the brain and the spinal cord (Fig. 3c, d).

Although these results indicate that CCR7 expression is sufficient to recruit leukaemic T-cells into the CNS, they do not address whether CCR7 alone is necessary for this function. We infected haematopoietic progenitors from $Ccr7^{-/-}$ (CCR7(KO))[17] and $Ccr7^{+/+}$ (wild-type) mice with Notch1-IC-expressing retroviruses (CCR7(KO)$^{Notch1-IC}$ and WT$^{Notch1-IC}$), and infected EGFP$^+$ cells were transplanted into wild-type hosts (WT/CCR7(KO)$^{Notch1-IC}$ and WT/WT$^{Notch1-IC}$). All recipient mice developed T-cell leukaemia, characterized by the detection of atypical peripheral blood CD4$^+$ CD8$^+$ cells as early as 2 weeks after transplant. As before, histopathological analysis showed notable tissue infiltration in all recipients (Supplementary Figs 2, 7 and data not shown). Hosts receiving leukaemic CCR7(KO) cells survived significantly longer than the hosts that received leukaemic wild-type cells (Supplementary Fig. 8). Although both hosts had similar leukaemic infiltration of most tissues, histopathological analysis of the CNS demonstrated that leukaemic CCR7(KO) cells were not found in the brain at 5 weeks after bone marrow transplantation, but leukaemic wild-type cells were (Fig. 4a). This was also true even at later time points just before the hosts succumbed, at 9–11 weeks after transplant (data not shown), uncoupling the tumour load to CNS infiltration. These results indicate that CCR7 is necessary for this process, and that the elimination of a single chemokine receptor *in vivo* is able to abrogate CNS involvement in T-ALL. CCR7 function seems to be specific for Notch1-induced T-cell malignancy, because deletion of this chemokine in two models of B-cell ALL failed to suppress CNS infiltration (Supplementary Fig. 9).

We then sought to exclude the possibility that CCR7-deficient cells do not infiltrate the CNS because of a more general defect in their ability to migrate. As discussed before, CCR7(KO)$^{Notch1-IC}$ cells infiltrate as well as leukaemic wild-type cells into several tissues. Also, the peripheral blood of hosts reconstituted with CCR7(KO)$^{Notch1-IC}$ contained

a similar (or even an increased) number of leukaemic blasts (Supplementary Figs 2 and 7). To attain a more dynamic view of the ability of the migratory properties of CCR7(KO)$^{Notch1-IC}$ and WT$^{Notch1-IC}$ leukaemic cells, we used two-photon imaging microscopy and traced EGFP$^+$ cells within the lymph nodes of living leukaemic hosts. As illustrated in Supplementary Movies 1 and 2, we were unable to demonstrate any differences in the ability of these cells to migrate within the lymph nodes. When individual cells were traced, no statistically significant differences were found between the velocity, turning angle, arrest coefficient, confinement and random walk[18] of CCR7(KO)$^{Notch1-IC}$ and WT$^{Notch1-IC}$ cells (Supplementary Fig. 10). These results underlined the specificity of CCR7 function in the targeting of T-ALL cells to the CNS.

Although these studies clearly demonstrate the importance of CCR7-mediated leukaemic cell recruitment to the CNS, they do not identify the ligand involved. It was previously shown that both chemokines are expressed in brain–blood barrier endothelium in an animal model of experimental autoimmune encephalomyelitis[19,20]. Moreover, CCL21 expression was detected previously in the choroid plexus—a possible site of lymphocyte entry into the subarachnoid space[21]. Thus, we performed immunohistochemical analyses to compare CCL19 and CCL21 expression in brain sections of control and leukaemic mice. CCL21 expression was undetectable in either sample (data not shown). However, CCL19 was detectable mainly in brain venules in the vicinity of infiltrating lymphocytes (Fig. 4b–g). Using immunofluorescence costaining, we confirmed that CD31$^+$ endothelial cells produced CCL19 (Fig. 4f, g). In addition, we purified brain CD31$^+$ endothelial cells and show a significant (8–10-fold when compared to CD31$^-$ brain cells) induction of $Ccl19$ message (Supplementary Fig. 11). Furthermore, endothelial cells from leukaemic animals expressed moderately higher (~twofold) levels of $Ccl19$ (Supplementary Fig. 12) when compared to endothelial brain cells from recipients that received EGFP-only infected cells (did not show any CNS infiltration). To prove that this slight overexpression is not, by itself, sufficient to attract CCR7$^+$ T cells in the CNS, we have transplanted wild-type CD4$^+$ T cells into leukaemic recipients. Analysis of recipient brain sections 12–16 h after transplantation failed to show any non-Notch1-IC–EGFP T-cell accumulation (Supplementary Fig. 12).

To prove the essential role of CCL19 expression, we modified our transplantation protocol and used *plt* mice as hosts, which lack CCL19 expression owing to a naturally occurring mutation[22]. We
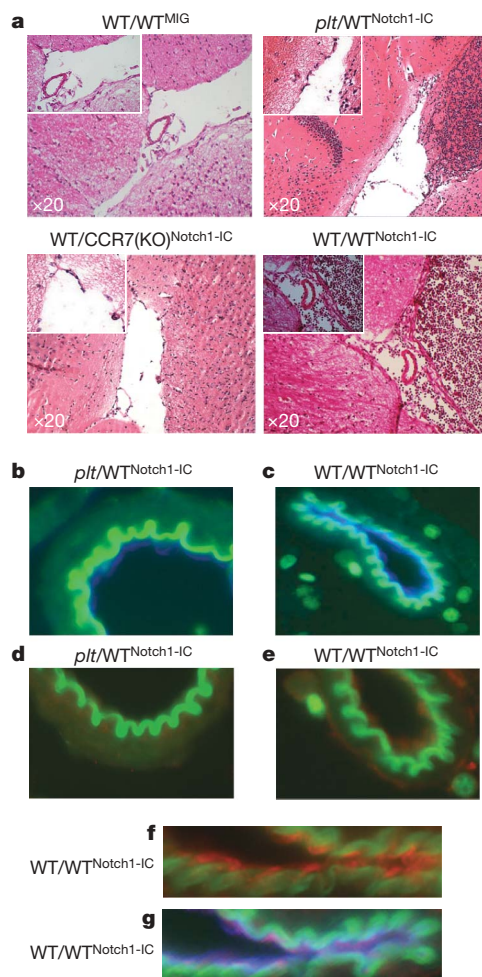
Figure 4 | CCR7–CCL19 interactions are essential for CNS infiltration in an animal model of T-ALL. a, Leukaemic cells in the brain meningeal spaces of the indicated transplanted recipients. A positive control (WT/WT[Notch1-IC]) is also included. **b–g,** Immunofluorescent staining of brain sections: magnification of a brain microvessel in *plt*/WT[Notch1-IC] (**b**) and WT/WT[Notch1-IC] (**c**) mice. Endothelial, CD31[+] cells are shown in blue. **d–g,** CCL19 (red) expression in WT/WT[Notch1-IC] (**e, f, g**) but not *plt*/WT[Notch1-IC] (**d**) microvessels. **f, g,** Magnification of a WT/WT[Notch1-IC] microvessel and costaining with CCL19 (red) and CD31 (blue).

transplanted these mice (or background- and age-matched controls) with Notch1-IC-expressing haematopoietic progenitors (*plt*/WT[Notch1-IC]). In agreement with our previous transplantations, *plt* mice survived longer than their wild-type counterparts (Supplementary Fig. 8). As before, there was an identical T-ALL induction and disease manifestation between different hosts, suggesting that the *plt* mutation does not affect leukaemic cell infiltration in peripheral lymphoid tissues (Supplementary Figs 2 and 7). However, leukaemic cells were unable to infiltrate the brain of *plt* hosts, further strengthening our argument that CCR7–CCL19 interactions are essential for CNS infiltration in T-ALL (Fig. 4a, b).

Our data demonstrated that a single chemokine–receptor pair is both necessary and sufficient for T-cell ALL leukaemic cell targeting to the CNS. Although CCR7 expression seems to be an important signal targeting T-ALL cells in the CNS, other factors that are present in transformed leukaemic cells but not in wild-type T cells are needed for CNS infiltration. These factors could include extra adhesion regulators (that is, integrins and metalloprotease, see Fig. 2a) that could potentially interact with CCR7 function. We believe that our studies open the way for the development of different therapeutic protocols, in which specific adhesion antagonists[23] can be used together with either current chemotherapy-based protocols or molecularly targeted

approaches, for example, the use of antagonists of Notch signalling[1]. Putative T-ALL CNS infiltration antagonists could include inhibitors of CCR7 expression and function, as well as drugs targeting specific migration regulators, which could be activated by CCR7 signalling.

METHODS SUMMARY

1. Grabher, C., von Boehmer, H. & Look, A. T. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nature Rev. Cancer* **6**, 347–359 (2006).
2. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature Rev. Immunol.* **8**, 380–390 (2008).
3. Pui, C. H. & Howard, S. C. Current management and challenges of malignant disease in the CNS in paediatric leukaemia. *Lancet Oncol.* **9**, 257–268 (2008).
4. Pui, C. H. & Evans, W. E. Treatment of acute lymphoblastic leukemia. *N. Engl. J. Med.* **354**, 166–178 (2006).
5. Cyster, J. G. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* **23**, 127–159 (2005).
6. Forster, R., Davalos-Misslitz, A. C. & Rot, A. CCR7 and its ligands: balancing immunity and tolerance. *Nature Rev. Immunol.* **8**, 362–371 (2008).
7. Weng, A. P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
8. Thompson, B. J. *et al.* The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. *J. Exp. Med.* **204**, 1825–1835 (2007).
9. Palomero, T. *et al.* Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nature Med.* **13**, 1203–1210 (2007).
10. O'Neil, J. *et al.* FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to γ-secretase inhibitors. *J. Exp. Med.* **204**, 1813–1824 (2007).
11. Vilimas, T. *et al.* Targeting the NF-κB signaling pathway in Notch1-induced T-cell leukemia. *Nature Med.* **13**, 70–77 (2007).
12. Klinakis, A. *et al.* Igf1r as a therapeutic target in a mouse model of basal-like breast cancer. *Proc. Natl Acad. Sci. USA* **106**, 2359–2364 (2009).
13. Charo, I. F. & Ransohoff, R. M. The many roles of chemokines and chemokine receptors in inflammation. *N. Engl. J. Med.* **354**, 610–621 (2006).
14. Cardona, A. E. L. i. M., Liu, L., Savarin, C. & Ransohoff, R. M. Chemokines in and out of the central nervous system: much more than chemotaxis and inflammation. *J. Leukoc. Biol.* **84**, 587–594 (2008).
15. Muller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
16. Shields, J. D. *et al.* Autologous chemotaxis as a mechanism of tumor cell homing to lymphatics via interstitial flow and autocrine CCR7 signaling. *Cancer Cell* **11**, 526–538 (2007).
17. Forster, R. *et al.* CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23–33 (1999).
18. Mempel, T. R., Scimone, M. L., Mora, J. R. & von Andrian, U. H. *In vivo* imaging of leukocyte trafficking in blood vessels and tissues. *Curr. Opin. Immunol.* **16**, 406–417 (2004).
19. Kivisakk, P. *et al.* Human cerebrospinal fluid central memory CD4[+] T cells: evidence for trafficking through choroid plexus and meninges via P-selectin. *Proc. Natl Acad. Sci. USA* **100**, 8389–8394 (2003).
20. Giunti, D. *et al.* Phenotypic and functional analysis of T cells homing into the CSF of subjects with inflammatory diseases of the CNS. *J. Leukoc. Biol.* **73**, 584–590 (2003).
21. Kivisakk, P. *et al.* Expression of CCR7 in multiple sclerosis: implications for CNS immunity. *Ann. Neurol.* **55**, 627–638 (2004).

22. Gunn, M. D. *et al.* Mice lacking expression of secondary lymphoid organ chemokine have defects in lymphocyte homing and dendritic cell localization. *J. Exp. Med.* **189,** 451–460 (1999).

23. Ransohoff, R. M. Natalizumab for multiple sclerosis. *N. Engl. J. Med.* **356,** 2622–2629 (2007).

24. Tseng, J. C. *et al.* Tumor-specific *in vivo* transfection with HSV-1 thymidine kinase gene using a Sindbis viral vector as a basis for prodrug ganciclovir activation and PET. *J. Nucl. Med.* **47,** 1136–1143 (2006).

25. Shakhar, G. *et al.* Stable T cell-dendritic cell interactions precede the development of both tolerance and immunity *in vivo*. *Nature Immunol.* **6,** 707–714 (2005).

26. Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427,** 154–159 (2004).

Author Contributions S.B., B.G.M. and I.A. designed experiments, performed experiments and wrote the manuscript. M.L.D., M.Li and M.G.R. performed the *in vivo* two-photon microscopy experiments and analysis. D.M. and J.-C.T. performed the bioluminescent imaging experiments. A.K. and A.E. generated the EF1-Notch1-IC mice. J.Z. performed the microarray data mining. Y.L., E.N. and D.Z. performed in the CNS pathology analysis. M.Lipp and J.S.B. assisted with experimental design, provided reagents and advice. T.T., L.R., S.C. and E.G. performed experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.A. (iannis.aifantis@nyumc.org).

nature

METHODS

Animals. C57BL/6 and $Rag2^{-/-} Il2rg^{-/-}$ mice were purchased from Jackson Laboratories and Taconic Farms. CCR7(KO) and *plt* mice have been described previously[17,22]. All mice were kept in specific pathogen-free animal facilities at the NYU-SOM and Mount Sinai Medical School. All animal procedures were performed in accordance to the guidelines of the Institutional Animal Care and Use Committee of the NYU-SOM.

Recombinant DNA constructs and retrovirus production and infection. The Notch1-IC retroviral plasmid, its parent CMMP-based vector[27], the pMXs IRES-mCherry retroviral vector and the pWPI lentivirus (a gift from E. Hernando) were used. Viral supernatants were generated as described previously[28]. Isolation, retroviral infection and reconstitution experiments were performed as previously described[27].

Antibodies and reagents. Mouse CD4-APC, CD8-PE-Cy7, CCR7-PE, B22 and CD31, and human CCR7-PE and CD3-APC primary antibodies were purchased from BD Bioscience. Human CD3 antibody was obtained by DAKO. Mouse CCL19 and CCL21 were procured from R&D. Mouse CD31 was from Pharmingen. Cytokines and chemokines were obtained from Peprotech. Secondary antibodies were purchased from Jackson Laboratories. Immunohistochemical analysis was performed using standard methods as described previously[11].

Chemotaxis assays. Chemotaxis assays were performed as previously described[29].

Quantitative RT–PCR. RNA was isolated using RNeasy Plus Mini Kit (Qiagen) columns and used to synthesize cDNA with the SuperScript First-Strand Kit (Invitrogen). Real-time PCR was performed using iQ SYBR Green Supermix and an iCycler (Bio-Rad). Relative expression was determined from cycle threshold (C_T) values, and normalized using β-actin as an internal control. All primer oligonucleotides are available on request.

Bioluminescent imaging. Imaging of luciferase-tagged leukaemic cells were performed as previously described[24].

Intravital microscopy. Two-photon imaging was performed as previously described[25]. Individual cell tracing and data analysis was performed as previously described[26].

Microarray analysis. The accession numbers for the individual array comparisons are Gene Expression Omnibus (GEO) series GSE6396, samples GSM147443, GSM147464 and GSM147508. Sample preparation and processing is detailed previously[11]. Pathway analysis of the microarray mRNA profiling results was performed using the Gene Ontology and KEGG pathway mapping within the web-based tool Database for Annotation, Visualization and Integrated Discovery. The results of category/pathway enrichment were manually curated for focused contents based on gene number and associated *P* value, and are summarized in Supplementary Table 1.

27. Sicinska, E. *et al.* Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer Cell* **4**, 451–461 (2003).

28. Ory, D. S., Neugeboren, B. A. & Mulligan, R. C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proc. Natl Acad. Sci. USA* **93**, 11400–11406 (1996).

29. Scimone, M. L., Aifantis, I., Apostolou, I., von Boehmer, H. & von Andrian, U. H. A multistep adhesion cascade for lymphoid progenitor cell homing to the thymus. *Proc. Natl Acad. Sci. USA* **103**, 7006–7011 (2006).

LETTERS

Genes that mediate breast cancer metastasis to the brain

Paula D. Bos[1], Xiang H.-F. Zhang[1], Cristina Nadal[1]†, Weiping Shu[1], Roger R. Gomis[1]†, Don X. Nguyen[1], Andy J. Minn[2], Marc J. van de Vijver[3], William L. Gerald[4], John A. Foekens[5] & Joan Massagué[1,6]

The molecular basis for breast cancer metastasis to the brain is largely unknown[1,2]. Brain relapse typically occurs years after the removal of a breast tumour[2–4], suggesting that disseminated cancer cells must acquire specialized functions to take over this organ. Here we show that breast cancer metastasis to the brain involves mediators of extravasation through non-fenestrated capillaries, complemented by specific enhancers of blood–brain barrier crossing and brain colonization. We isolated cells that preferentially infiltrate the brain from patients with advanced disease. Gene expression analysis of these cells and of clinical samples, coupled with functional analysis, identified the cyclooxygenase COX2 (also known as PTGS2), the epidermal growth factor receptor (EGFR) ligand HBEGF, and the α2,6-sialyltransferase ST6GALNAC5 as mediators of cancer cell passage through the blood–brain barrier. EGFR ligands and COX2 were previously linked to breast cancer infiltration of the lungs, but not the bones or liver[5,6], suggesting a sharing of these mediators in cerebral and pulmonary metastases. In contrast, ST6GALNAC5 specifically mediates brain metastasis. Normally restricted to the brain[7], the expression of ST6GALNAC5 in breast cancer cells enhances their adhesion to brain endothelial cells and their passage through the blood–brain barrier. This co-option of a brain sialyltransferase highlights the role of cell-surface glycosylation in organ-specific metastatic interactions.

Brain metastasis affects an estimated 10% of cancer patients with disseminated disease[2,8,9]. Even small lesions can cause neurological disability, and the median survival time of patients with brain metastasis is short. The two main sources of brain metastasis—adenocarcinomas of the lung or the breast—represent different models of the course of the disease. Metastasis from lung adenocarcinomas develops within months of diagnosis and affects several organs besides the brain[10]. This course suggests that aggressive pro-metastatic functions foster the colonization of several organs at once. In breast cancer, a long period of remission often precedes distant relapse[3,4], suggesting that breast cancer cells initially lack the full competence for outgrowth in distant organs but develop this under the selective pressure of different organ microenvironments. Breast cancer metastasis frequently becomes prevalent in one organ long before it does in others, and brain metastasis tends to be a late event[2]. The barriers to metastasis are distinct in different organs. Capillary endothelia are backed by a basement membrane in the lung[11] and also by tight junctions and astrocyte foot processes in the blood–brain barrier (BBB)[2,8], whereas the capillaries in the bone marrow and the liver are fenestrated[11,12]. The composition of the parenchyma also varies extensively between these organs. The protracted progression of disseminated cancer cells in different environments may give rise to metastatic speciation, as suggested by the coexistence of malignant cells with different organ tropisms in fluids from patients with advanced disease[5,13]. Analysis of such malignant cell populations has revealed genes that selectively mediate breast cancer metastasis to bones[13] or the lungs[5]. Here we adopted this approach to test the hypothesis that breast cancer infiltration of the brain requires general mediators of extravasation, complemented by specific enhancers of cell passage through the BBB.

We used oestrogen-receptor-negative (ER$^-$) pleural malignant cells from a Memorial Sloan-Kettering Cancer Center (MSKCC) breast cancer patient (CN34 sample), and also MDA-MB-231 cells (MDA231 for brevity)—an ER$^-$ breast cancer pleural cell line previously used for the isolation of bone and lung metastatic cells[5,13] and brain metastatic cells[14]. CN34 and MDA231 cells were inoculated into the arterial circulation of immunodeficient female mice to isolate populations that target the brain (Fig. 1a–d). After tumour dissociation and expansion in culture, the resulting cell populations (brain metastatic derivative 1, BrM1) were subjected to a second round of *in vivo* selection, yielding BrM2 cell populations that showed a significant increase in brain metastatic activity (Fig. 1a, b). When grown as mammary tumours, CN34-BrM2 metastasized to brain in 42% (5 out of 12) of the mice, whereas parental CN34 mammary tumours yielded no brain metastases in ten mice. BrM2 cells showed no increase in bone or lung metastatic activity compared to the parental populations (Supplementary Table 1). MDA231 lung metastatic (LM2-4175) and bone metastatic (BoM-1833; refs 5, 13) derivatives were poorly metastatic to brain compared to BrM2 cells (Fig. 1b). The CN34-BrM2 and MDA231-BrM2 cell lines generated multifocal lesions in the cerebrum, the cerebellum and the brainstem (Fig. 1e, f and Supplementary Fig. 1a, b), and in the leptomeninges (Fig. 1g and Supplementary Fig. 1c, d). Larger nodules developed hemorrhagic cores and oedema (Fig. 1d). Astrogliosis occurred in the periphery of the tumours (Fig. 1h). All of these features are typical of brain metastasis in breast cancer patients[2,15]. Within 24 h of inoculation, BrM2 cells lodged in brain capillaries as single cells (Fig. 1i), suggesting that brain metastases resulted from an ability of these cells to breach the BBB.

Comparative genome-wide expression analysis demonstrated 243 genes that were overexpressed or underexpressed in the brain metastatic populations of both cell lines, or were upregulated in one cell system and overexpressed in the other, or were downregulated in one system and underexpressed in the other (Supplementary Table 2). To prioritize these candidate genes, we screened for those whose expression in breast tumours was associated with brain relapse. Univariate

[1]Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [2]Department of Radiation and Cellular Oncology, and Ludwig Center for Metastasis Research, The University of Chicago, Chicago, Illinois 60637, USA. [3]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [4]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [5]Department of Medical Oncology, Erasmus MC Rotterdam, Josephine Nefkens Institute and Cancer Genomics Centre, Rotterdam, The Netherlands. [6]Howard Hughes Medical Institute, Chevy Chase, Maryland 20185, USA. †Present addresses: Institut de Malalties Hemato-Oncològiques, Hospital Clínic, 08036 Barcelona, Spain (C.N.); Oncology Programme, Institute for Research in Biomedicine, 08028 Barcelona, Spain (R.R.G.).
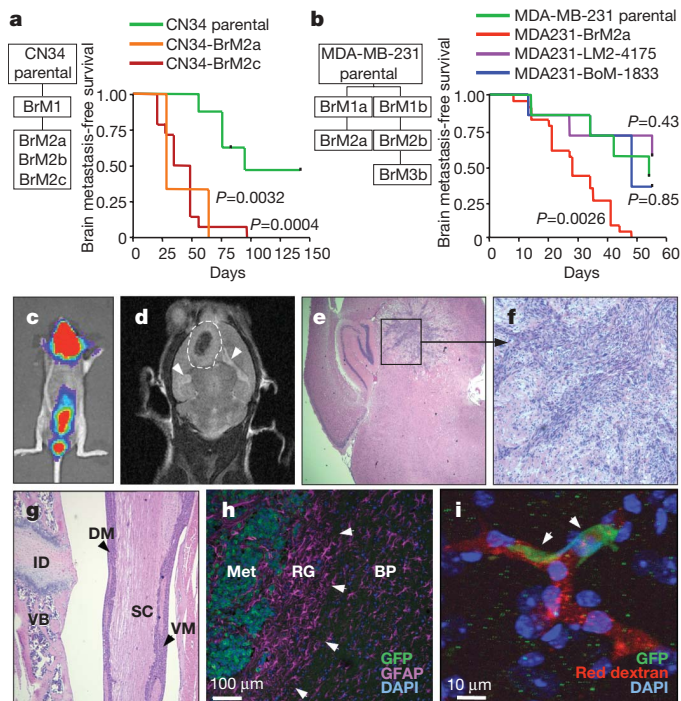
Figure 1 | Isolation and characterization of brain metastatic variants.
a, **b**, Flowcharts of the *in-vivo*-selected brain metastatic derivatives, and Kaplan–Meier survival curves for brain metastasis-free survival of representative CN34 (parental $n = 8$, BrM2c $n = 14$, BrM2a $n = 3$) (**a**) and MDA231 (parental $n = 7$, LM $n = 7$, BoM $n = 7$, BrM2a $n = 23$) (**b**) cell line variants. A log-rank test was used to compare the survival curves of each cell line to the parental line. BoM, BrM and LM indicate bone, brain and lung metastatic derivative, respectively. **c**, Bioluminescence image of a mouse with brain and leptomeningeal metastasis by CN34-BrM2c cells. **d**, Magnetic resonance imaging (MRI) of a brain metastatic lesion (dashed line) showing a hemorrhagic core, and brain oedema (arrowheads). **e**, **f**, Representative haematoxylin and eosin (H&E)-stained sections of a mouse brain containing a CN34-BrM2c lesion (original magnification, ×2 (**e**) and ×10 (**f**)). **g**, H&E staining of a section showing MDA231-BrM2a cell colonization of the dorsal (DM) and ventral (VM) meninges. ID, intervertebral disc; SC, spinal cord; VB, vertebral body. Original magnification, ×5. **h**, MDA231-BrM2a brain metastatic lesion showing reactive glia (RG, arrowheads) around the metastatic lesion (Met). Tumour cells express green fluorescent protein (GFP), and glial cells are stained with the glial marker glial fibrillary protein (GFAP, purple). BP, brain parenchyma; DAPI, 4,6-diamidino-2-phenylindole. **i**, GPF$^+$ MDA231-BrM2a cells (arrowheads) arrested in brain capillaries (red, rhodamine dextran) 24 h after intracardiac injection into mice. Nuclei were stained with DAPI (blue).

analysis in a combined cohort of 368 clinically annotated breast tumours (MSK-82 and EMC-286 sets; Supplementary Table 3) showed 17 genes whose expression was correlated ($P < 0.05$) with brain relapse (Supplementary Table 4), and resembled the expression profile in the brain-metastatic-derived (BrM) cells (Supplementary Fig. 2a, b). A classifier trained with this brain metastasis gene set (BrMS) showed association with brain relapse in two independent breast tumour data sets (Fig. 2a and Supplementary Fig. 3a). The same procedures applied to randomly generated sets of 500 genes yielded no classifiers that performed in the data sets. The association of BrMS status with brain relapse remained significant within ER$^-$ tumours (Supplementary Table 5 and Supplementary Fig. 3b), and was stronger in patients who received no adjuvant therapy ($P < 0.0001$; Fig. 2b and Supplementary Table 6). BrMS$^+$ tumours appeared in different molecular subtypes of breast cancer[16] (Supplementary Fig. 4a–c).

We do not interpret the results of the gene expression analysis as reflecting the only possible 17 genes associated with brain relapse. However, the expression of these 17 genes in breast tumours was not
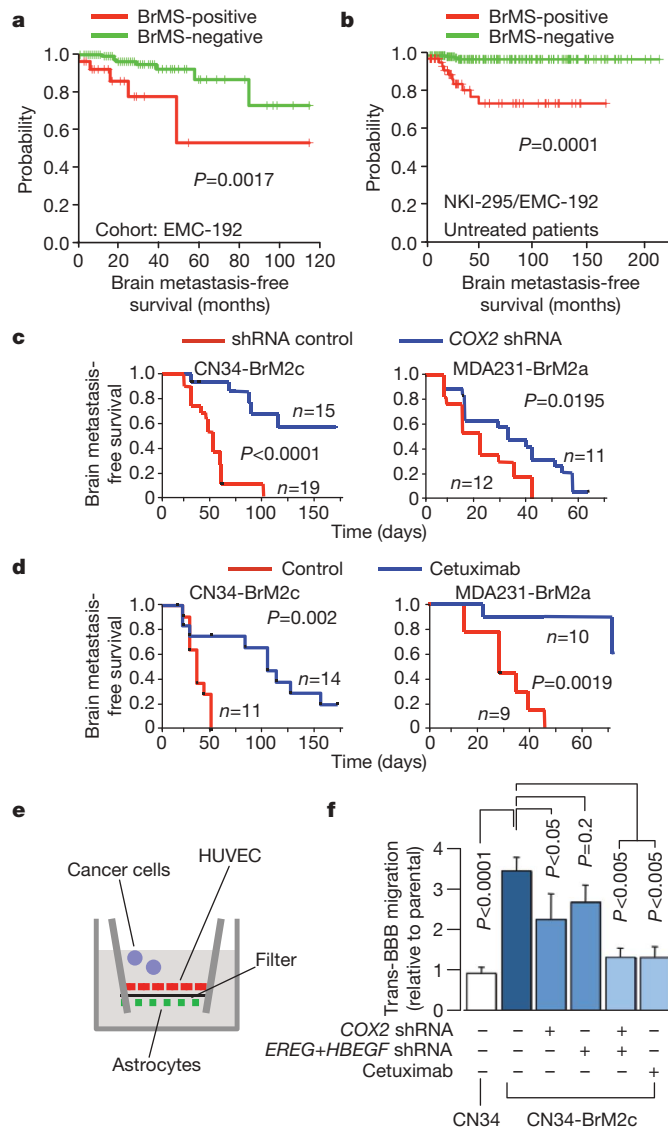

Figure 2 | COX2 and EGFR ligands as mediators of brain metastasis and BBB transmigration. **a**, **b**, Kaplan–Meier curves for brain metastasis-free survival on the basis of BrMS status in an independent cohort of 192 breast tumours (**a**), and in a combined cohort of 262 breast tumours from patients who received no adjuvant therapy (**b**). **c**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with the indicated cell lines expressing short hairpin RNA (shRNA) vector control or shRNA targeting *COX2*. **d**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with the indicated cell lines and treated with cetuximab or vehicle control. **e**, Schematic of the *in vitro* BBB model assay system. HUVEC, human umbilical vein endothelial cells. **f**, *In vitro* BBB transmigration activity of the indicated cell lines and conditions. The number of transmigrated cells relative to the parental cell lines is plotted. Error bars, s.e.m.; $n = 6$–20. P values were determined by log rank test (**a**–**d**) and one-tailed unpaired *t*-test (**f**).

associated with relapse to bones, liver or lymph nodes (Supplementary Fig. 5a). Notably, six of these genes were shared with an 18-gene lung metastasis signature (LMS) that is associated with relapse to the lungs, but not to bones, liver or lymph nodes[17]. LMS$^+$ status was weakly associated with relapse to the brain, and BrMS$^+$ status with relapse to the lungs (Supplementary Fig. 5b–d). The shared genes include the prostaglandin-synthesizing enzyme cyclooxygenase-2 (*COX2*), which promotes extravasation in the lungs[6]; collagenase-1 (*MMP1*), which mediates invasion and extravasation[6,18]; angiopoietin-like 4 (*ANGPTL4*), which is induced by tumour-derived TGF-β and disrupts endothelial junctions[19]; latent TGF-β-binding protein (*LTBP1*), which controls TGF-β activation[20]; fascin-1 (*FSCN1*), which supports cancer

cell migration[21]; and the putative metastasis suppressor *RARRES3* (ref. 5). Furthermore, both gene sets include an EGFR ligand: heparin-binding EGF (*HBEGF*) in the BrMS, and epiregulin (*EREG*) in the LMS. *EREG* was highly expressed in CN34-BrM but not in MDA231-BrM cells.

These observations suggested a partial sharing of mediators of metastasis to the brain and lungs, a hypothesis that we tested by focusing on EGFR ligands and COX2. Prostaglandin production during inflammation increases BBB permeability[22]. HBEGF induces cancer cell motility and invasiveness[23]. The brain metastatic activity of BrM2 cells (Fig. 2c, d) was decreased by RNA interference (RNAi)-mediated knockdown of *COX2* expression[6] (Supplementary Fig. 6a, b), or by treatment of mice with cetuximab, which targets human EGFR[24]. To investigate cancer cell passage through the BBB, we used an *in vitro* model consisting of human primary endothelial cells and astrocytes (Fig. 2e and Supplementary Fig. 7a). This model generates a tight barrier that expresses brain endothelial markers and lacks permeability to albumin (Supplementary Fig. 7b–d)[25]. CN34-BrM2 and MDA231-BrM2 cells were three- to fourfold more active than their parental lines at migrating through this barrier. *COX2* knockdown inhibited this transmigration in both BrM2 lines, as did the addition of cetuximab to CN34-BrM2 cells, or the combination of *COX2*, *HBEGF* and *EREG* knockdowns (Fig. 2f and Supplementary Fig. 7e).

The ability of COX2 and EGFR ligands to prime breast cancer cells for extravasation into the brain may explain the association of lung and brain relapse in breast cancer[9,26]. However, given the course of breast cancer, we postulated that brain relapse also depends on selective mediators of infiltration through the unique barriers of the brain. As candidates we selected genes whose expression was increased more than threefold in all CN34-BrM2 and MDA231-BrM2 isolates, but not in bone metastatic[13] or lung metastatic MDA231 derivatives[5]. After excluding histone genes, we arrived at a set of 26 candidates (Supplementary Table 7). This set largely consists of cell–cell interaction components. The α2,6-sialyltransferase *ST6GALNAC5* stood out because its expression is normally restricted to the brain both in mice[7] and in humans (Supplementary Fig. 8). *ST6GALNAC5* messenger RNA levels were notably higher in brain metastatic derivatives than in parental cell lines in MDA231 samples (30-fold ± 1 higher; mean ± s.d.), CN34 cells (>100-fold), and in two other pleural-derived samples that were subjected to one cycle of selection for brain infiltration in mice (CN37-BrM1, 95-fold ± 23; CN41-BrM1, 72-fold ± 12).

Sambucus nigra agglutinin (SNA), a lectin that binds to α2,6-linked sialyl groups, specifically stained mammary tumours and brain lesions formed by BrM2 cells (Fig. 3a, b and Supplementary Fig. 9a, b). Fifty per cent (6 out of 12) of brain metastatic samples from breast cancer patients had areas with strong SNA staining, whereas 18% (2 out of 11) of lung metastatic samples had areas of low SNA staining, and the remaining 9 samples were negative (Fig. 3c, d). Furthermore, in a set of ER⁻ breast cancer metastasis to various sites, *ST6GALNAC5* mRNA levels roughly equalled that of the BrM2 cell lines in 23% (3 out of 12) of brain metastases, but not in metastases to other sites (Fig. 3e; $P = 0.04$, Fisher's exact test).

Sialyltransferases catalyse the addition of sialic acid to gangliosides and glycoproteins[27], and cell-surface sialylation has been implicated in cell–cell interactions[28]. CN34-BrM2 cells were more adhesive to monolayers of human primary brain endothelial cells than were the CN34 or *ST6GALNAC5*-knockdown CN34-BrM2 cells (Supplementary Fig. 10a). Notably, the knockdown of *ST6GALNAC5* decreased the BBB transmigration activity of CN34-BrM2 cells to ground level (Fig. 4a), and also decreased the brain metastatic activity of CN34-BrM2 cells (Fig. 4b). Brain metastasis was further decreased by combination with cetuximab treatment (Fig. 4b). *ST6GALNAC5* knockdown did not inhibit the growth of CN34-BrM2c cells in culture or as mammary tumours, the basal lung-seeding ability of these cells, or the aggressive lung-colonizing ability of the MDA231 derivative
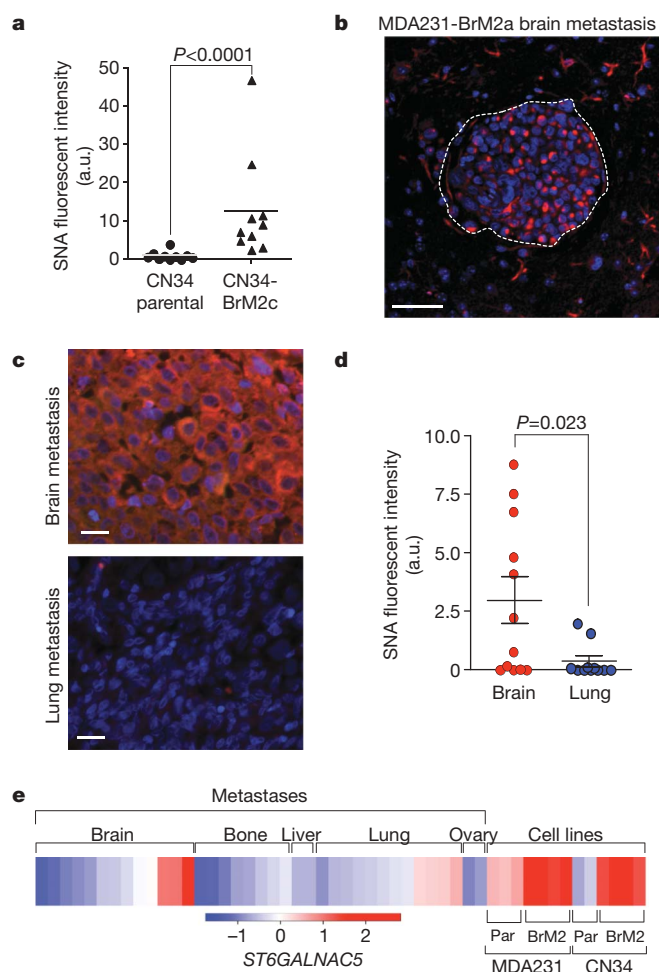

Figure 3 | *ST6GALNAC5* expression and activity in brain metastasis from breast cancer. a, Quantification of SNA staining in mammary fat pad tumours formed by parental CN34 or CN34-BrM2c cells in mice. a.u., arbitrary units. **b**, SNA staining of a mouse brain metastasis after intracardiac inoculation of MDA231-BrM2a cells. Scale bar, 50 μm **c**, SNA staining of representative human brain and lung metastases samples from the same breast cancer patient. Scale bars, 20 μm. **d**, Distribution of SNA staining intensity, quantified by Metamorph analysis, in 12 brain and 11 lung metastases resected from breast cancer patients. *P* values (**a, d**) were determined by Mann–Whitney one-tailed test. **e**, Heat map showing the relative *ST6GALNAC5* expression levels in a panel of 13 brain, 8 bone, 3 liver, 12 lung and 2 ovary human metastases from breast cancer patients. Included for comparison are the parental (par) and brain metastatic derivatives from MDA231 and CN34 cells. Data are on the basis of Affymetrix probe intensity.

line LM2-4175 (ref. 5) (Supplementary Fig. 10b–f). However, transduction of LM2-4175 cells with an *ST6GALNAC5* expression vector increased the ability of these cells to transmigrate across a BBB (Fig. 4c) and to infiltrate the brain (Supplementary Fig. 11a). Infiltrated brains showed a prevalence of micrometastases in mice inoculated with LM2-ST6 and of larger lesions in mice inoculated with BrM2 cells (Fig. 4d and Supplementary Fig. 11b). ST6GALNAC5 therefore mediates infiltration into the brain, and further mediators may be required for the expansion of the resulting foci into macrometastases.

Breast cancer cells can disseminate to the lungs from early stages of tumour development[29], indicating that cancer cells departing from breast tumours are competent for extravasation through lung microcapillary walls. Our results indicate that the expression of COX2 and HBEGF in primary tumours enhances cancer cells for extravasation through the non-fenestrated capillaries of the brain and lungs, whereas ST6GALNAC5 expression is co-selected with, and acts as a specific mediator of, cancer cell infiltration through the blood–brain
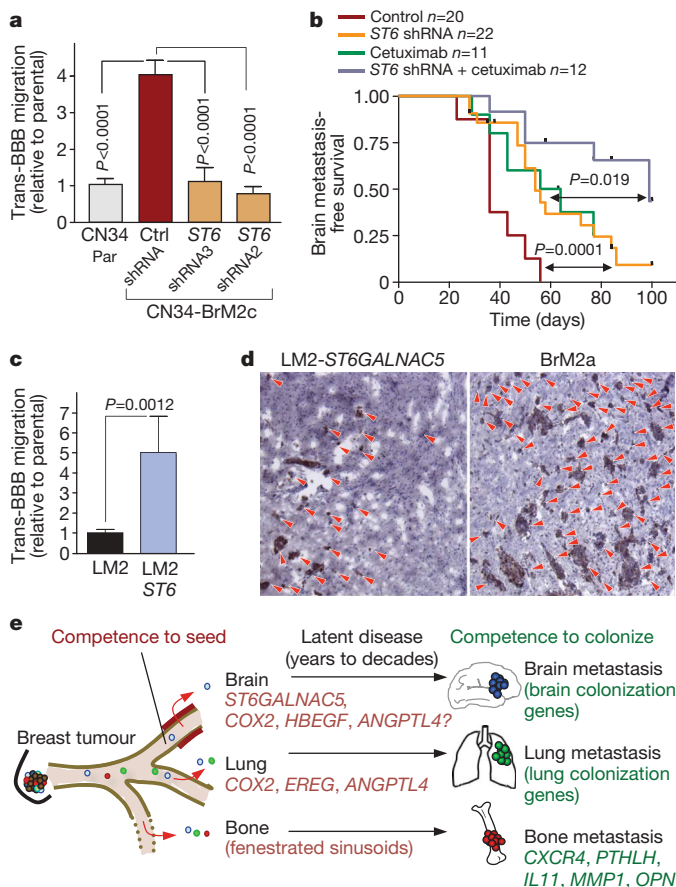
Figure 4 | The brain-specific sialyltransferase *ST6GALNAC5* as a BBB extravasation and brain metastasis gene. a, *In vitro* BBB transmigration activity of the indicated cell lines. Ctrl, control; par, parental; *ST6*, *ST6GALNAC5*. Error bars, s.e.m.; $n = 9$–27; *P* values were determined by Mann–Whitney one-tailed test. **b**, Kaplan–Meier curves for brain metastasis-free survival of mice injected with CN34-BrM2 cells expressing an shRNA targeting *ST6GALNAC5* or an empty vector control, and then treated with cetuximab or vehicle. *P* values were determined using log-rank test. **c**, *In vitro* BBB transmigration activity of LM2 cells transduced with an empty vector or with *ST6GALNAC5*. Error bars, s.e.m.; $n = 22$–27; *P* values were determined by Mann–Whitney one-tailed test. **d**, Anti-GFP immunostaining of representative lesions from mice injected intracardially with the indicated cell lines. Red arrowheads show individual tumour foci; original magnification, ×10. **e**, Schematic model of organ-specific metastatic extravasation of breast cancer cells. Extravasation into the bone marrow is a relatively permissive process owing to the fenestrated endothelium lining the sinusoid capillaries. Extravasation into the pulmonary or brain parenchyma requires specific functions for breaching the non-fenestrated capillary walls of these organs. Shared mediators of extravasation include, among others, *COX2* and EGFR ligands such as epiregulin and *HBEGF*. Passage through the BBB requires further mediators including, but not limited to, the brain-specific sialyltransferase *ST6GALNAC5*. Competence to colonize each organ requires additional mediators.

barrier (Fig. 4e). These findings draw attention to the role of cell-surface sialylation as a previously unrecognized participant in brain metastasis, and to the possibility of therapeutically disrupting these interactions. Our work also points to other candidate brain metastasis genes, including genes implicated in vascular permeability and leukocyte infiltration during brain inflammatory processes, and genes implicated in neurite extension and astrocyte cell processes. The role of these genes in brain metastasis and their interest as therapeutic targets is open to further analysis.

METHODS SUMMARY

CN34 tumour cells were isolated from the pleural effusion of a breast cancer patient treated at our institution, after written consent in accordance with

Institutional Review Board (IRB) regulations. Brain metastatic populations from these cells and MDA-MB-231 cells were obtained by consecutive rounds of *in vivo* selection in 6–7-week-old beige nude and athymic mice, respectively. All animal work was done in accordance with the MSKCC Institutional Animal Care and Use Committee. Methods for RNA extraction, labelling and hybridization for DNA microarray analysis have been described previously[17]. Bioinformatics analyses with detailed descriptions can be found in the Methods. Knockdown and overexpression of candidate genes, and cetuximab inhibitor studies were performed as previously described[6]. The *in vitro* BBB model was set up as previously described[25], and modified to enable tumour cell counting. *Sambucus nigra* lectin staining was performed using standard histochemical techniques, and quantified using Metamorph software analysis. The Methods section provides further information, including malignant cell isolation from pleural fluids, tumour cell extraction and cell culture protocols, animal inoculation and bioluminescence imaging, generation of retroviral gene knockdown and overexpression vectors, transfections and infections, RNA and protein expression, *in vitro* BBB transmigration assay, endothelial cell adhesion assay, and metastatic tissue staining and quantification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Chiang, A. C. & Massague, J. Molecular basis of metastasis. *N. Engl. J. Med.* **359**, 2814–2823 (2008).
2. Weil, R. J. *et al.* Breast cancer metastasis to the central nervous system. *Am. J. Pathol.* **167**, 913–920 (2005).
3. Karrison, T. G., Ferguson, D. J. & Meier, P. Dormancy of mammary carcinoma after mastectomy. *J. Natl. Cancer Inst.* **91**, 80–85 (1999).
4. Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl Acad. Sci. USA* **100**, 7737–7742 (2003).
5. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
6. Gupta, G. P. *et al.* Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* **446**, 765–770 (2007).
7. Okajima, T. *et al.* Molecular cloning of brain-specific GD1α synthase (ST6GalNAc V) containing CAG/glutamine repeats. *J. Biol. Chem.* **274**, 30557–30562 (1999).
8. El Kamar, F. G. & Posner, J. B. Brain metastases. *Semin. Neurol.* **24**, 347–362 (2004).
9. Lassman, A. B. & DeAngelis, L. M. Brain metastases. *Neurol. Clin.* **21**, 1–23 (2003).
10. Feld, R., Rubinstein, L. V. & Weisenberger, T. H. Sites of recurrence in resected stage I non-small-cell lung cancer: a guide for future studies. *J. Clin. Oncol.* **2**, 1352–1358 (1984).
11. Inoue, S. & Osmond, D. G. Basement membrane of mouse bone marrow sinusoids shows distinctive structure and proteoglycan composition: a high resolution ultrastructural study. *Anat. Rec.* **264**, 294–304 (2001).
12. Paku, S., Dome, B., Toth, R. & Timar, J. Organ-specificity of the extravasation process: an ultrastructural study. *Clin. Exp. Metastasis* **18**, 481–492 (2000).
13. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
14. Yoneda, T. *et al.* A bone-seeking clone exhibits different biological properties from the MDA-MB-231 parental human breast cancer cells and a brain-seeking clone *in vivo* and *in vitro*. *J. Bone Miner. Res.* **16**, 1486–1495 (2001).
15. Fitzgerald, D. P. *et al.* Reactive glia are recruited by highly proliferative brain metastases of breast cancer and promote tumor cell colonization. *Clin. Exp. Metastasis* **25**, 799–810 (2008).
16. Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
17. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl Acad. Sci. USA* **104**, 6740–6745 (2007).
18. Egeblad, M. & Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nature Rev. Cancer* **2**, 161–174 (2002).
19. Padua, D. *et al.* TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**, 66–77 (2008).
20. Saharinen, J., Hyytiainen, M., Taipale, J. & Keski-Oja, J. Latent transforming growth factor-β binding proteins (LTBPs)–structural extracellular matrix proteins for targeting TGF-β action. *Cytokine Growth Factor Rev.* **10**, 99–117 (1999).
21. Adams, J. C. Roles of fascin in cell adhesion and motility. *Curr. Opin. Cell Biol.* **16**, 590–596 (2004).
22. de Vries, H. E. *et al.* The influence of cytokines on the integrity of the blood-brain barrier *in vitro*. *J. Neuroimmunol.* **64**, 37–43 (1996).
23. Miyamoto, S. *et al.* Heparin-binding epidermal growth factor-like growth factor as a novel targeting molecule for cancer therapy. *Cancer Sci.* **97**, 341–347 (2006).
24. Goldstein, N. I. *et al.* Biological efficacy of a chimeric antibody to the epidermal growth factor receptor in a human tumor xenograft model. *Clin. Cancer Res.* **1**, 1311–1318 (1995).

25. Eugenin, E. A. & Berman, J. W. Chemokine-dependent mechanisms of leukocyte trafficking across a model of the blood-brain barrier. *Methods* **29**, 351–361 (2003).
26. Slimane, K. *et al.* Risk factors for brain relapse in patients with metastatic breast cancer. *Ann. Oncol.* **15**, 1640–1644 (2004).
27. Harduin-Lepers, A. *et al.* The human sialyltransferase family. *Biochimie* **83**, 727–737 (2001).
28. Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconj. J.* **18**, 841–850 (2001).
29. Husemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58–68 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions P.D.B. and J.M. designed experiments, analysed data and wrote the manuscript. J.M. supervised the research. X.H.-F.Z. performed bioinformatics analyses. P.D.B. performed experiments. W.S. assisted with experiments. C.N. and R.R.G. isolated metastatic cells from clinical samples. D.X.N. helped with gliosis immunostaining and confocal microscopy. A.J.M. identified LMS clinical correlation with brain relapse. W.L.G., J.A.F. and M.J.V.d.V. obtained, classified and processed breast tumour samples. All authors discussed the results and commented on the manuscript.

Author Information The clinical microarray data on the brain metastatic cell lines have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under the GEO series accession number GSE12237. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M. (j-massague@ski.mskcc.org).

METHODS

Isolation of carcinoma cells from pleural effusions. Clinical specimens were obtained from three consenting patients (CN34, CN37 and CN41) with metastatic breast cancer treated at our institution, following IRB-approved protocols. Epithelial cells were obtained from pleural fluids as described before[30]. In brief, pleural fluid was collected in the presence of heparin (5 U ml^{-1}), and centrifuged at 150g for 10 min. Cell pellets were resuspended in PBS, red blood cells were lysed with ACK lysis buffer, and a fraction of the cells was subjected to negative selection to remove leukocytes (CD45$^+$ and CD15$^+$ populations). Cells were cultured for 24 h, and epithelial cells were sorted from this population using an anti-EpCAM antibody. The resulting cell population was transduced with a lentivirus expressing the triple-fusion reporter encoding herpes simplex virus thymidine kinase 1, GFP and firefly luciferase[31]. GFP-expressing cells were sorted and maintained at 5% CO$_2$ at 37 °C in M199 medium supplemented with 2.5% fetal bovine serum (FBS), 10 µg ml^{-1} insulin, 0.5 µg ml^{-1} hydrocortisone, 20 ng ml^{-1} EGF, 100 ng ml^{-1} cholera toxin, 1 µg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin, for approximately 1 week before mouse injection.

Isolation of brain metastatic cells. A cell suspension containing 10^5 CN34 breast cancer cells in a volume of 100 µl was injected in the left cardiac ventricle of anesthetized 6–7-week-old Cr:NIH-bg-nu-Xid mice. A cell suspension of 10^4 MDA-MB-231 breast cancer cells in a volume of 100 µl was injected in the left cardiac ventricle of anesthetized 6–7-week-old athymic mice. Tumour development was monitored by weekly bioluminescence imaging using the IVIS-200 imaging system from Xenogen as previously described[5]. Brain metastatic lesions were confirmed by MRI and histological analysis after necropsy. Brain lesions were localized by *ex vivo* bioluminescence imaging, and resected under sterile conditions. Half of the tissue was fixed with 4% paraformaldehyde (PFA), and processed for histological analysis. The other half was minced and placed in culture medium containing a 1:1 mixture of DMEM/Ham's F12 supplemented with 0.125% collagenase III and 0.1% hyaluronidase. Samples were incubated at room temperature for 4–5 h, with gentle rocking. After collagenase treatment, cells were briefly centrifuged, resuspended in 0.25% trypsin, and incubated for a further 15 min in a 37 °C water bath. Cells were resuspended in culture medium and allowed to grow to confluence on a 10-cm dish. GFP$^+$ cells were sorted for further propagation in culture or inoculation in mice. All animal work was done following a protocol approved by the MSKCC Institutional Animal Care and Use Committee.

Histological analysis and microscopy. Brain metastatic lesions were fixed with 4% PFA overnight, washed twice with PBS, dehydrated in 50% ethanol, and subsequent 70% ethanol, and embedded in paraffin for H&E staining. For all other purposes, animals were perfused with 10 ml PBS, and pre-fixed with 5 ml of 4% PFA. Lesions were extracted and post-fixed with 4% PFA for a further 2 h, incubated in a solution of 30% sucrose in PBS for 1–2 days, and processed for OCT compound embedding and montage. Assessment of reactive glia was performed by staining with the astrocyte marker GFAP (DAKO), followed by detection with a fluorescently labelled secondary antibody. Detection of tumour foci in the brain parenchyma was achieved by immunohistochemistry using an anti-GFP antibody (Invitrogen). For this purpose, the brain hemispheres were separated longitudinally and mounted in OCT. Four sections from each hemisphere were obtained, separated by 300 µm. In total, eight brain regions were stained with anti-GFP and microscopically explored using ×5 magnification. Microscopic analysis was performed using a Zeiss Axioplan2 microscope, and foci quantification was achieved using Metamorph software analysis. For detection of tumour cells in the brain microvasculature, 10^6 brain metastatic cells were injected into the left cardiac ventricle of anesthetized mice. Enhancement of the green fluorescence was obtained by labelling the tumour cells with 5 µM CFMDA cell-tracker dye (Invitrogen) for 45 min before injection. To visualize the brain vasculature, mice were injected with 2 mg g^{-1} of body weight rhodamine-labelled 70-kDa dextran (Invitrogen) via retro-orbital inoculation 1 h before culling. Animals were perfused and culled 24 h after tumour cell inoculation, and brain was processed for OCT compound embedding. Thirty-micrometre sections were examined on an Upright Leica TCS SP2 confocal microscope, and ×63 images were collected.

RNA isolation and gene-expression profiling. RNA was extracted from exponentially growing cells using the RNeasy mini kit (Qiagen). Labelling and hybridization of the samples to HG-U133A gene expression chip (Affymetrix) were performed by the MSKCC Genomics Core Facility using standard methodology, as previously described[17].

Data analysis was performed using the GeneSpring 7.2 software. The raw data was filtered by intensity values equal to or greater than 150. Class comparison between parental and brain metastatic populations (three biological replicates of the MDA231 parental versus four brain metastatic derivatives, and two biological replicates of CN34 parental versus four brain metastatic derivatives) was performed to identify gene expression changes of 2.5-fold associated with the brain metastatic phenotype ($P < 0.05$). This step yielded 271 genes (310 probe sets) that were differentially expressed between the parental and the brain-metastatic CN34 cell lines, and 179 genes (210 probe sets) between the parental and the brain-metastatic MDA231 cell lines. From these lists, we selected genes with a concordant expression change of at least 1.5-fold in the other cell system, genes that were upregulated in one system and consistently expressed at a high level in the other, and genes that were downregulated in one system and consistently underexpressed in the other. As a result, we obtained a final list of 280 probe sets, corresponding to 243 genes that are associated with brain metastatic phenotype in CN34 and MDA231 cells (Supplementary Table 2).

BrMS derivation and clinical sample analysis. Microarray data from four cohorts of breast tumours were used for analysis. The MSK-82 cohort was more locally advanced compared to either the NKI-295 or the EMC-286 series (91% T2-T4 and 66% node positive in MSK-82, compared to 47% T2-T4 and 49% node positive in NKI-295, and 51% T1, 46% T2 and 0% node positive in EMC-286). EMC-192 is a heterogeneous cohort that includes 144 tumours from patients that relapsed and received first-line chemotherapy for metastatic disease in an adjuvant or recurrence setting, and 48 from patients that were node-negative and did not receive adjuvant systemic therapy. Sites of recurrence and treatment information for each of these data sets are shown in Supplementary Table 3.

The MSK-82 and EMC-286 cohorts were analysed on Affymetrix HG-U133A platform, and the EMC-192 cohort on HG-U133 plus 2.0. The NKI-295 set was analysed on Agilent microarrays. To achieve statistical power given the limited incidence of brain metastasis in these cohorts, we merged the MSK-82 and EMC-286 cohorts. All data sets were first transformed to log-2 scales and median-centred. Z transformation was then performed to normalize gene expression across all samples in each cohort, using the MAS5.0 normalization approach[32].

The 243 gene set (280 probe set) associated with brain metastasis was used to fit a Cox hazard ratio regression model to gauge the association of each gene with brain or lung metastasis-free survival in the combined MSK-82/EMC-286 cohort. This was achieved using the survival package in the R statistical software. Wald test was used to calculate the P values. We designated 'brain metastasis gene set' (BrMS) the 17 genes with $P < 0.05$ from the 243 gene set. To examine whether these 17 genes were obtained owing to artefactual effects caused by merging the MSK-82 and EMC-286 cohorts, we performed principal component analysis and tested whether the two data sets differentially distributed along the top two principal components (Supplementary Fig. 12). This analysis revealed no segregation between the two data sets ($P = 0.11$ and 0.28 for the top two principal components by t-test) (Supplementary Fig. 9), arguing that the statistical significance and classification power of the 17 genes cannot be attributed to the merge of the two data sets. The identification of BrMS$^+$ tumours was achieved by unsupervised hierarchical clustering of tumours in the MSK-82 and EMC-286 as a training cohort. The resulting cluster-tree was cut at different distance cutoffs to yield different numbers (2 to 10) of sub-clusters. In each case, the cluster that most resembled the gene expression pattern of BrM cells was compared with the other clusters for enrichment of brain relapse events, using Fisher's exact test. The best cutoff was determined when such cluster not only maintained the resemblance of gene expression pattern to BrM cells, but also best-segregated brain relapse events. This cluster was defined as BrMS$^+$. Heat maps were generated using the gplots package of R statistical program.

For each patient, metastasis-free survival is defined as the time interval between the surgery and the diagnosis of metastasis. When there are several metastatic sites, because the chronological order of different metastatic sites on the same patient was not annotated, it has been assumed that all metastases were diagnosed at the same time, and each patient was scored for each of the corresponding sites in our analyses.

BrMS$^+$ and BrMS$^-$ tumours were used to train a support vector machine (package e1071, R statistical program). We used a linear kernel and expression values of the 17-gene set as features. The trained classifier was then applied to the NKI-295 and EMC-192 cohorts to predict BrMS$^+$ tumours. We performed Kaplan–Meier analysis and log-rank tests on the survival rates of the predicted BrMS$^+$ and BrMS$^-$ tumours in the NKI-295 and EMC-192 data sets, using the survival package of R. The specificity and sensitivity of the BrMS and LMS classifiers as predictors of brain metastatic recurrence is shown in Supplementary Table 6.

As a control, we randomly generated ten sets of 500 genes each, and determined the correlation of the genes in each set with brain metastasis-free survival in the 368-tumour cohort. We then selected univariately significant genes, trained classifiers using EMC/MSK-368 and tested the classifiers on the independent EMC192 and NKI295 data sets as we did with the BrMS. Such randomly derived classifiers yielded P values for brain relapse association ranging from 0.1

to 0.8 (data not shown). Thus, the performance of the BrMS as a correlate of brain relapse was not due to chance.

Knockdown and overexpression cell lines. Knockdown of *COX2* and *EREG* with a validated hairpin was achieved as previously described[6]. Knockdown of *HBEGF* was achieved with pRetroSuper vector targeting the sequence 5′- GGTA TGCTGTCATGGTCCT-3′, and knockdown of *ST6GALNAC5* was achieved by targeting the sequences 5′-CATAAGCAACTCAACAATA-3′ (shRNA2), and 5′-GAGCACATCTCCACTGACT-3′ (shRNA3). Overexpression of *ST6GALNAC5* was achieved by cloning the open reading frame (ORF) of this gene into the pBabe-Puro retroviral vector from complementary DNA obtained from normal human brain using the primers 5′-GGAATTCATGTACCCATACGATGTTC CAGATTACGCTAAGACCCTGATGCGCCA-3′ and 5′- ACGCGTCGACTTAGA ACACAGGTTTATTCT-3′.

The efficiency of the knockdown and overexpression was confirmed by quantitative PCR with reverse transcription (qRT–PCR) TaqMan gene expression assays (Applied Biosystems), or western immunoblotting analysis (anti-COX2 antibody, Cayman Chemical). β2-microglobulin and actin were used as endogenous controls for qRT–PCR and western blot, respectively. The viral particles for infection of the brain-metastatic derivatives were obtained by transfection of the GPG29 amphotropic packaging cell line, and collection of supernatants 48 and 72 h after transfection. Supernatants were filtered and centrifuged at 43,000*g* to concentrate the viral particles, and were used to infect sub-confluent cultures in the presence of 5 μg ml^{-1} polybrene overnight. Puromycin (2 μg ml^{-1}) was used to select for stable cell lines. Only cell lines with a transduction rate over 80–90% were used for further studies.

Cetuximab treatment. Biweekly intraperitoneal injection of 1 mg cetuximab antibody (ImClone) was performed as previously described[6]. Animals were given one or two doses of cetuximab before intracardiac inoculation of the tumour cells, and were maintained on drug treatment until the end of the experiment.

***In vitro* blood–brain barrier assay.** Primary human umbilical vein endothelial cells (HUVEC, ScienCell) were co-cultured with human primary astrocytes (ScienCell), on opposite sides of a polylysine-treated, gelatin-coated tissue culture transwell insert for 3 days as previously described[25]. In brief, 3 μm pore PET tissue culture inserts (Fisher) were treated with polylysine (1 μg ml^{-1}, Millipore) overnight, washed four times, and coated with 0.2% gelatin (Sigma) for a minimum of 30 min. Inserts were placed upside-down in a 15 cm plate, and 10^5 primary human astrocytes were plated on the membrane surface. Astrocytes were fed every 15 min for 5 h, and the inserts were then flipped and placed in 24-well plates. Fifty-thousand endothelial cells were plated on the upper chamber of the inserts, and cultures were placed in the incubator, without further perturbation. Thee days later, the tightness of the barriers was tested by permeability to serum albumin. Evans blue-conjugated albumin (0.45% in phenol red-free medium) was added to the upper chamber and incubated for 30 min at 37 °C. Medium from the bottom chamber was collected, and absorbance was measured at 620 nm. Controls included astrocytes alone, endothelial cells alone, astrocytes plated on both sides of the insert, and insert alone. Immunofluorescence staining was performed with antibodies against the endothelial cell marker von Willebrand factor (Sigma), the tight junction marker zonula occludens 1 (Zymed), and the astrocyte marker GFAP (Dako). The expression of the brain endothelial markers glucose transporter-1 (*GLUT1*, also known as *SLC2A1*) and γ-glutamyl transpeptidase (*GGT1*) was evaluated by qPCR using TaqMan gene expression assays (Applied Biosystems) after RNA extraction from the top cell layer of the BBB inserts.

For BBB transmigration assays, cancer cells were labelled with 5 μM CFMDA cell tracker green (Invitrogen) for 45 min, and recovered overnight before assaying. Fifty-thousand cells were seeded on the upper chamber and incubated for 14–18 h. Inserts were washed with PBS and fixed with 4% PFA for 20 min. The membranes were removed from the plastic insert and mounted on a microscope slide. Pictures from 5–8 inserts per experiment were taken, and the number of transmigrated cells was counted. For cetuximab experiments, tumour cells were pre-treated with 10 μg ml^{-1} cetuximab in the medium for 5 h, and the same concentration was added to the tumour cell suspension seeded on the top of the endothelial layer.

Lectin staining. SNA staining was performed on perfused, paraffin-embedded xenograft tumour tissue as previously described[33]. In brief, after standard deparafinization, sections were washed with PBS, and endogenous peroxidase was quenched by incubation in 0.3% H_2O_2 in methanol for 30 min at room temperature. Sections were washed three times with PBS and blocked in 10% donkey serum for 30 min at room temperature. Labelling with biotin-conjugated SNA was carried out at a concentration of 100 μg ml^{-1} for 45 min, followed by three washes with PBS. An Alexa-568 conjugated-tyramide amplification kit (Invitrogen) was used following manufacturer's procedures to detect the biotinylated lectin. Sections were mounted with Prolong Gold mounting medium (Invitrogen), and images were taken using a Zeiss Axioplan2 microscope. The

same protocol was followed for SNA staining of human breast cancer metastatic tissues, except that SNA was used at 10 μg ml^{-1}. Metamorph analysis was used to quantify the intensity of the lectin staining, and the resulting values were plotted in a scatter plot graph for comparison.

Metastatic sample gene-expression analysis. Human breast carcinoma metastasis specimens were obtained from the files of the Department of Pathology in compliance with protocols approved by the MSKCC IRB. Samples were snap-frozen in liquid nitrogen and stored at −80 °C. Each sample was examined histologically using H&E-stained cryostat sections. Regions were manually dissected from the frozen block to provide consistent tumour cell content of greater than 70% in tissues used for analysis. All studies were conducted under MSKCC IRB-approved protocols. RNA was extracted from frozen tissues by homogenization in Trizol reagent (GIBCO/BRL) and evaluated for integrity. cDNA was synthesized from total RNA using a T7-promoter-tagged-dT primer. RNA target was synthesized by *in vitro* transcription and labelled with biotinylated nucleotides (Enzo Biochem). Labelled target was assessed by hybridization to Test3 arrays (Affymetrix). All gene expression analysis was carried out using HG-U133A GeneChip. Gene expression was quantified using MAS 5.0 or GCOS (Affymetrix).

Oestrogen receptor status of the metastatic samples was confirmed by examination of the probe '205225_at' that represents the oestrogen receptor gene *ESR1*. Only metastases with raw intensity values <1,000 were kept for further analyses. For *ST6GALNAC5* expression, we first set the median expression value of all probes on each chip to be zero, and carried out a *Z*-transformation across different samples. The *ST6GALNAC5* expression level was then interrogated by the intensity of the probe '220979_s_at', and a heat map was made accordingly with the previously described tools.

Oncomine gene expression data analysis. Relative levels of *ST6GALNAC5* mRNA expression in human tissues were obtained by Oncomine Cancer Microarray database analysis (http://www.oncomine.org)[34] of a published gene expression data set[35]. The data was log-2-transformed, with the media set to zero and s.d. set to one.

Cell adhesion assay. Primary human brain microvascular endothelial cells (hBMVECs, ScienCell) were grown to confluency in 12-well plates. Before seeding the tumour cells, hBMVEC monolayers were washed twice with 0.5% BSA in PBS. Tumour cells were briefly trypsinized, resuspended in medium containing 0.5% BSA, and counted. Five-hundred-thousand cells were plated in each well, and allowed to adhere to the monolayer for 30 min. Plates were washed three times for 5 min each, shaking. Cells were lysed with 100 μl Passive lysis buffer (Promega) for 1 h, shaking. Firefly luciferase activity was determined using an Orion microplate luminometer (Berthold Detection Systems). Assays were performed in quadruplicate.

Other tissue culture procedures. Primary human endothelial cells and astrocytes were cultured in M199 medium supplemented with 50 mg ml^{-1} ascorbic acid, 25 mg ml^{-1} heparin, 3 mg ml^{-1} endothelial cell growth supplement (Sigma), 5 μg ml^{-1} bovine brain extract (Clonetics), 20% FBS, 5% human serum (Biocell), 1 μg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin. GPG29 cells were cultured in DMEM supplemented with 20 ng ml^{-1} doxycycline, 2 μg ml^{-1} puromycin, 0.3 mg ml^{-1} G418, and 10% FBS. 293T/17 packaging cell lines used for lentiviral production, and MDA-MB-231 parental cell lines and derivatives were cultured in DMEM supplemented with 10% FBS, 1 μg ml^{-1} fungizone, and 100 U ml^{-1} penicillin/streptomycin. All transfections were performed using Lipofectamine2000 (Invitrogen). GPG29 cells were maintained in DMEM supplemented with 10% FBS and 1 mM sodium pyruvate after transfection. *In vitro* proliferation quantification was achieved by plating 10^4 cells in each of 24-well plates, and lysing them in triplicates every other day with 100 μl with Passive lysis buffer (Promega) for 1 h, shaking. Firefly luciferase activity was determined using an Orion microplate luminometer (Berthold Detection Systems).

Other animal procedures. Orthotopic tumour growth was measured by injecting 1 × 10^6 viable single cells in a 1:1 mixture of PBS and growth-factor-reduced Matrigel (BD Biosciences) into mammary gland 4 in a total volume of 50 μl as previously described[5]. Primary tumour growth rates were analysed by measuring tumour length (*L*) and width (*W*), and calculating tumour volume on the basis of the formula $\pi LW^2/6$. For experimental lung metastasis assays, 2 × 10^5 cells were resuspended in 0.1 ml PBS and injected into the lateral tail vein. Lung metastatic progression was monitored and quantified using non-invasive bioluminescence as previously described[5].

30. Gomis, R. R. *et al.* C/EBPβ at the core of the TGF-β cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* 10, 203–214 (2006).
31. Ponomarev, V. *et al.* A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* 31, 740–751 (2004).
32. Cheadle, C., Cho-Chung, Y. S., Becker, K. G. & Vawter, M. P. Application of z-score transformation to Affymetrix data. *Appl. Bioinformatics* 2, 209–217 (2003).

nature

33. Kaneko, Y., Yamamoto, H., Colley, K. J. & Moskal, J. R. Expression of Galβ1,4GlcNAc α2,6-sialyltransferase and α2,6-linked sialoglycoconjugates in normal human and rat tissues. *J. Histochem. Cytochem.* **43**, 945–954 (1995).

34. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).

35. Shyamsundar, R. *et al.* A DNA microarray survey of gene expression in normal human tissues. *Genome Biol.* **6**, R22 (2005).

LETTERS

The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs

Michele Trabucchi[1], Paola Briata[2]*, MariaFlor Garcia-Mayoral[3], Astrid D. Haase[4], Witold Filipowicz[4], Andres Ramos[3], Roberto Gherzi[2]* & Michael G. Rosenfeld[1]*

Consistent with the role of microRNAs (miRNAs) in down-regulating gene expression by reducing the translation and/or stability of target messenger RNAs[1], the levels of specific miRNAs are important for correct embryonic development and have been linked to several forms of cancer[2–4]. However, the regulatory mechanisms by which primary miRNAs (pri-miRNAs) are processed first to precursor miRNAs (pre-miRNAs) and then to mature miRNAs by the multiprotein Drosha and Dicer complexes[5–8], respectively, remain largely unknown. The KH-type splicing regulatory protein (KSRP, also known as KHSRP) interacts with single-strand AU-rich-element-containing mRNAs and is a key mediator of mRNA decay[9,10]. Here we show in mammalian cells that KSRP also serves as a component of both Drosha and Dicer complexes and regulates the biogenesis of a subset of miRNAs. KSRP binds with high affinity to the terminal loop of the target miRNA precursors and promotes their maturation. This mechanism is required for specific changes in target mRNA expression that affect specific biological programs, including proliferation, apoptosis and differentiation. These findings reveal an unexpected mechanism that links KSRP to the machinery regulating maturation of a cohort of miRNAs that, in addition to its role in promoting mRNA decay, independently serves to integrate specific regulatory programs of protein expression.

We analysed the immunopurified Dicer-containing complex[11] by mass spectroscopy and identified, amongst other proteins, KSRP (Supplementary Fig. 1a)—a highly conserved nucleo-cytoplasmic RNA-binding protein regulating distinct steps of mRNA life cycle[12,13] (Supplementary Fig. 1b–d). KH domains 3 and 4 of KSRP (KH3–4) are required to promote AU-rich element (ARE)-containing labile mRNA decay[9,10]. Coimmunoprecipitation revealed that KSRP is an integral component of the Dicer complex in HeLa cells (Fig. 1a). On Dicer-knockdown-induced (Supplementary Fig. 2a) pre-miRNA upregulation, an anti-KSRP antibody immunoprecipitated pre-let-7a[14] (Fig. 1b). Recombinant KSRP directly interacted with pre-let-7a-1, and KH3–4 accounted for the high-affinity binding to RNA (Supplementary Fig. 3a). KSRP interacted with the terminal loop (TL) of pre-let-7a-1 (TL-let-7a-1) but did not associate with either single- or double-stranded mature let-7a (Fig. 1c), with KH3–4 accounting for KSRP binding to TL-let-7a-1 (Supplementary Fig. 3b).

We titrated the protein with increasing amounts of TL-let-7a-1, and TL-let-7a-1 with increasing amount of protein, while monitoring the binding by NMR and circular dichroism, respectively. KH3–4 binds to TL-let-7a-1 with a 1:1 stoichiometry and a dissociation constant (K_d) of ~50 nM (Supplementary Fig. 3c) whereas single KH3 and KH4 domains bind to TL-let-7a-1 with a K_d of ~500 nM and ~40 μM, respectively (Supplementary Fig. 3d, e). In contrast to
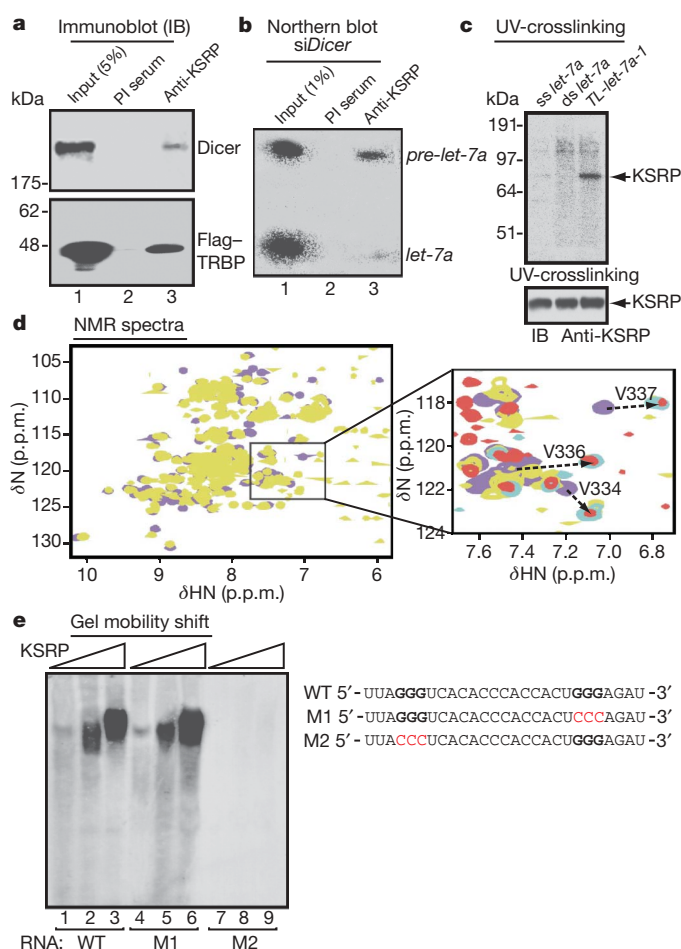

Figure 1 | KSRP, a component of Dicer complex, interacts with the TL of pre-let-7a-1. a, Coimmunoprecipitation of endogenous KSRP and either Dicer or Flag–TRBP in HeLa cell extracts. PI serum, preimmune serum. **b,** Coimmunoprecipitation of KSRP and pre-let-7a in HeLa cells transiently transfected with Dicer small interfering RNA (siDicer). **c,** KSRP (300 nM) binds to TL-let-7a-1 but not to single-stranded (ss) or double-stranded (ds) let-7a. UV, ultraviolet. **d,** Superimposition of ^{15}N-^{1}H HSQC spectra of KH3–4-free (violet) and bound to TL-let-7a-1 (yellow). In the magnified panel, spectra of bound KH3 (cyan) and KH2-3 (red) are also shown. Arrows highlight the shift of a few representative peaks in the core of the RNA-binding groove. **e,** Interaction of KSRP (50–300 nM) with either wild-type (WT) TL-let-7a-1 or two distinct mutants (M1 and M2). Bold indicates G stretches and red indicates mutations.

[1]Howard Hughes Medical Institute, Department and School of Medicine, University of California, San Diego, 9500 Gilman Drive, Room 345, La Jolla, California 92093-0648, USA. [2]Istituto Nazionale per la Ricerca sul Cancro (IST), Largo R. Benzi, 10; 16132 Genova, Italy. [3]Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. [4]Friedrich Miescher Institute for Biomedical Research, PO Box 2543, 4002 Basel, Switzerland.
*These authors contributed equally to this work.

what was observed in the KH3–TNF-α ARE interaction[10], KH3 recognizes a specific site in the *TL-let-7a-1* and contributes most of the binding affinity in the KH3–4–RNA interaction, whereas KH4 has an auxiliary role. Comparison of the chemical shift changes undergone by KH3 amide resonances upon RNA binding in the isolated KH3 and within the two-domain KH2–3 and KH3–4 constructs showed that the bound position of the resonance affected by *TL-let-7a-1* binding is the same whether or not KH2 or KH4 is present (Fig. 1d and Supplementary Fig. 3f). NMR spectra showed that the position of nearly all of the resonances of the RNA-bound KH3 is the same for the *TL-let-7a-1* and the entire *pre-let-7a-1* (Supplementary Fig. 3g).

Our recent work indicated that KH3 recognizes short G-rich stretches with high specificity and affinity[15]. The *TL-let-7a-1* presents two GGG triplets, supporting the idea that KH3 docks KSRP on a specific site and that KSRP–*pre-let-7a-1* recognition takes place using a very different mode and a considerably higher affinity compared with ARE mRNA targets[10,16]. Mutational analysis of *TL-let-7a-1* revealed that the 5′ GGG triplet accounts for high-affinity binding to KSRP (Fig. 1e).

KSRP knockdown in both HeLa and NIH-3T3 cells (Supplementary Fig. 2b, c) abrogated the endogenous let-7a-mediated post-transcriptional silencing of a reporter construct containing six let-7a binding

sites (let-7a6×BS) (Fig. 2a). *KSRP* knockdown inhibited the effect of both transfected *pri-* and *pre-let-7a-1* on let-7a6×BS reporter but left transfected mature let-7a function unaffected (Fig. 2b and Supplementary Fig. 4a, b). Control *HNRNPD* (previously known as *AUF1*) knockdown (Supplementary Fig. 2d) had no effect (Fig. 2a, b).

Recombinant KSRP increased the processing activity of Dicer (Fig. 2c) whereas KSRP immunodepletion from 293T extracts (Supplementary Fig. 2e) removed the *pre-let-7a-1* processing activity (Supplementary Fig. 4c) leaving mature let-7a unaffected (Supplementary Fig. 4d). Finally, immunopurified KSRP-containing complexes specifically processed *pre-let-7a-1* into mature let-7a (Supplementary Fig. 4e–g).

To investigate whether all pre-miRNAs are regulated by KSRP, we performed miRNA microarray analysis. Transient *KSRP* knockdown in HeLa cells significantly reduced (>1.5-fold) the expression of 14 miRNAs (Supplementary Fig. 5a) and reduced by 1.2–1.5-fold the expression of 20 additional miRNAs (Supplementary Table 1). Northern blot analysis in both HeLa and NIH-3T3 cells confirmed that let-7a, miR-26b, miR-20, miR-106a, miR-21 and miR-16 were reduced by 40–70% upon *KSRP* knockdown whereas miR-23b and miR-24 were unaffected (Fig. 2d, Supplementary Fig. 5b–e and data not shown). Ultraviolet crosslinking and *in vitro* processing experiments showed the selectivity of KSRP binding and KSRP-induced
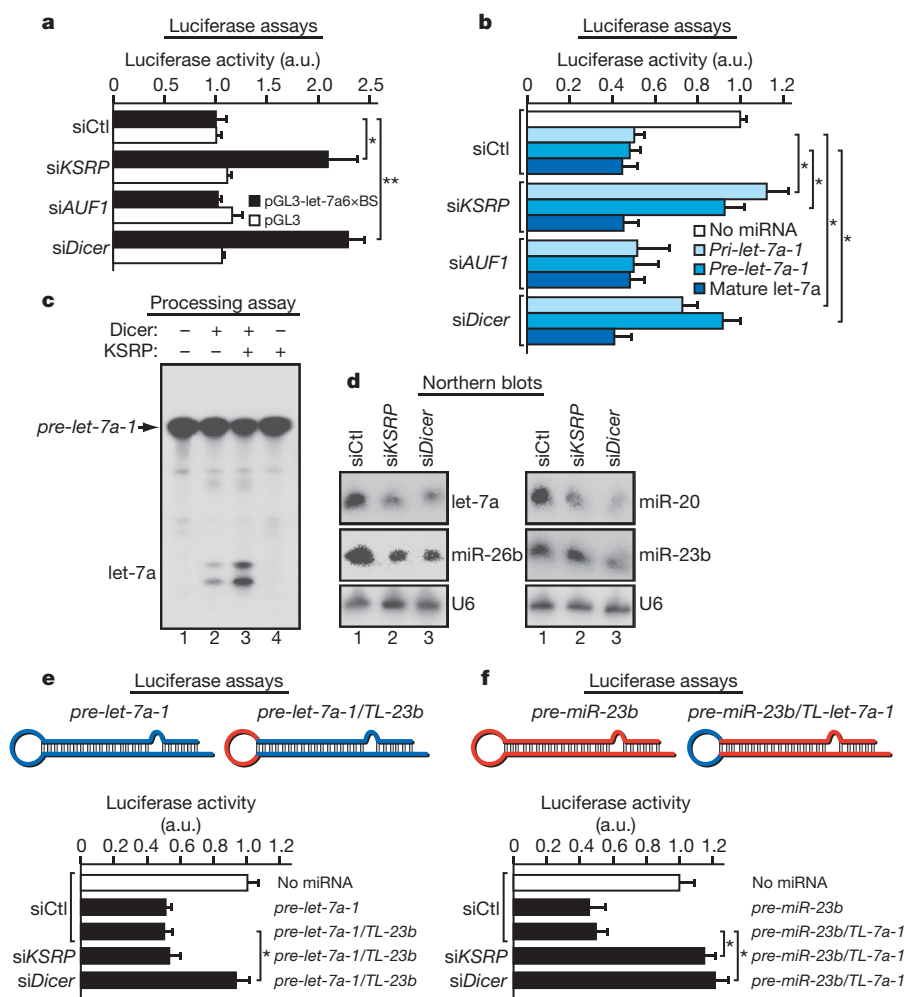

Figure 2 | KSRP regulates *pre-let-7a-1* processing and controls the expression of certain miRNAs. a, *KSRP* knockdown reduces the effect of endogenous let-7a on the activity of pGL3-let-7a6×BS. Student's *t*-test: *P < 0.05, **P < 0.01. a.u., arbitrary units; siCtl, control siRNA. **b,** *KSRP* knockdown reduces the effect of either *pri-let-7a-1* or *pre-let-7a-1*, but not of mature let-7a, on pGL3-let-7a6×BS activity. *P < 0.05. **c,** Recombinant KSRP increases the processing activity of recombinant Dicer on *pre-let-7a*.

d, Analysis of total RNA from control, *KSRP*- or *Dicer*-knockdown HeLa cells. **e, f,** *KSRP* knockdown does not reduce the effect of a chimaeric *pre-let-7a-1*, comprising the TL of *pre-miR-23b*, on the activity of pGL3-let-7a6×BS (**e**) but impairs the effect of a chimaeric *pre-miR-23b*, comprising the TL of *pre-let-7a-1*, on the activity of pGL3-miR-23b3×BS (**f**). *P < 0.05. All data are presented as mean and s.d. (*n* = 4).

1011

processing for those miRNAs for which expression was regulated by KSRP (Supplementary Fig. 6a–c and data not shown). Interestingly, the *TL-miR-21* does not contain any GGG triplets but displays two potential non-optimal binding sites for KSRP KH3 and KH4 (Supplementary Fig. 6d and ref. 15). Mutation of the two G residues in the GUUG 5′ element abrogated the KSRP interaction whereas mutation of the 3′ GG doublet only reduced the binding affinity (Supplementary Fig. 6d).

To investigate the function of TL–KSRP interaction, we used chimaeric pre-miRNAs with swapped loop sequences. *KSRP* knockdown did not affect the expression of a reporter containing let-7a-binding sites when a chimaeric *pre-let-7a-1* comprising the TL of miR-23b was expressed in HeLa cells (Fig. 2e). Conversely, expression of a reporter containing miR-23b-binding sites was impaired by *KSRP* knockdown in HeLa cells expressing a chimaeric *pre-miR-23b* containing the TL of let-7a-1 (Fig. 2f and Supplementary Fig. 6e).

Because pre-miRNA accumulation induced by *KSRP* knockdown was low in comparison to that induced by *Dicer* knockdown (Supplementary Fig. 7a), and *KSRP* knockdown increased the levels of *pri-let-7a-1* and *pri-miR-21* (Supplementary Fig. 7b), we hypothesized an involvement of KSRP in pri-miRNA processing. KSRP coimmunoprecipitated with Flag-tagged Drosha and DGCR8 (Fig. 3a), and anti-KSRP antibody immunoprecipitated *pri-let-7a-1* and *pri-miR-21* but not *pri-miR23b*, *pri-miR-24* and *pri-miR-17* (Fig. 3b and Supplementary Fig. 7c). Endogenous, transfected and recombinant KSRP specifically interacted with *pri-let-7a-1* (Supplementary Fig. 8a–c). Immunopurified KSRP-containing complexes processed *pri-let-7a-1* similarly to Drosha-containing immunopurified complexes (Supplementary Fig. 9a). Either stable or transient *KSRP* knockdown in 293T, HeLa and NIH-3T3 cells reduced the *pri-let-7a-1* processing (Supplementary Fig. 9b–d) leaving the processing of *pri-miR-23b* unaffected (Supplementary Fig. 9b,

right panel, and data not shown). Addition of recombinant KSRP to 293T cells stably transfected with short hairpin RNA to *KSRP* (293T sh*KSRP*) extracts restored *pri-let-7a-1* processing whereas KSRP overexpression in 293T cells strongly increased *pri-let-7a-1* processing (Fig. 3c and Supplementary Fig. 9e).

We explored the possibility that KSRP favours the association of enzymatic complexes with certain miRNA precursors. Indeed, *KSRP* knockdown abrogated the interaction of Drosha with *pri-let-7a-1* and *pri-miR-21* (Fig. 3d) and strongly reduced the binding of Dicer to *pre-let-7a-1* and *pre-miR-21* (Fig. 3e). In contrast, *KSRP* knockdown did not affect the interaction of the same pri-miRNAs and pre-miRNAs with either DGCR8 or TARBP2P (also known as TRBP; Supplementary Fig. 10a and data not shown)[5,11].

KSRP knockdown increased mRNA levels of two let-7 targets, NRAS and MYC[17,18] (Supplementary Fig. 10b), and specificity was established because cotransfection of mature let-7a abolished this effect (Supplementary Fig. 10c, d). Furthermore, *KSRP* knockdown in U2OS osteosarcoma cells reduced the expression of mature let-7a, significantly upregulated cell proliferation[19] (Fig. 4a and data not shown) and reduced the anti-proliferative effect of transfected *pri-let-7a-1* but not of mature let-7a (Fig. 4a and Supplementary Fig. 11a). Similarly, *KSRP* knockdown prevented *pri-miR-16-1*-induced apoptosis[20] but did not affect the activity of transfected mature miR-16 (Supplementary Fig. 11b, c). Recently, an essential role of certain miRNAs (miR-1, miR-133a and miR-206) in C2C12 myoblast differentiation has been reported[21,22]. *KSRP* knockdown in C2C12 reduced the maturation of 'myogenic' miRNAs (Supplementary Fig. 12a–c). The interaction of KSRP with *pri-miR-206*, *pri-miR-1-1* and *pri-miR-1-2* was increased by pro-differentiative stimuli (differentiation medium, Supplementary Fig. 12d). Finally, *KSRP* knockdown inhibited the miR-206-induced downregulation of direct target mRNAs, including those encoding gap junction protein α1
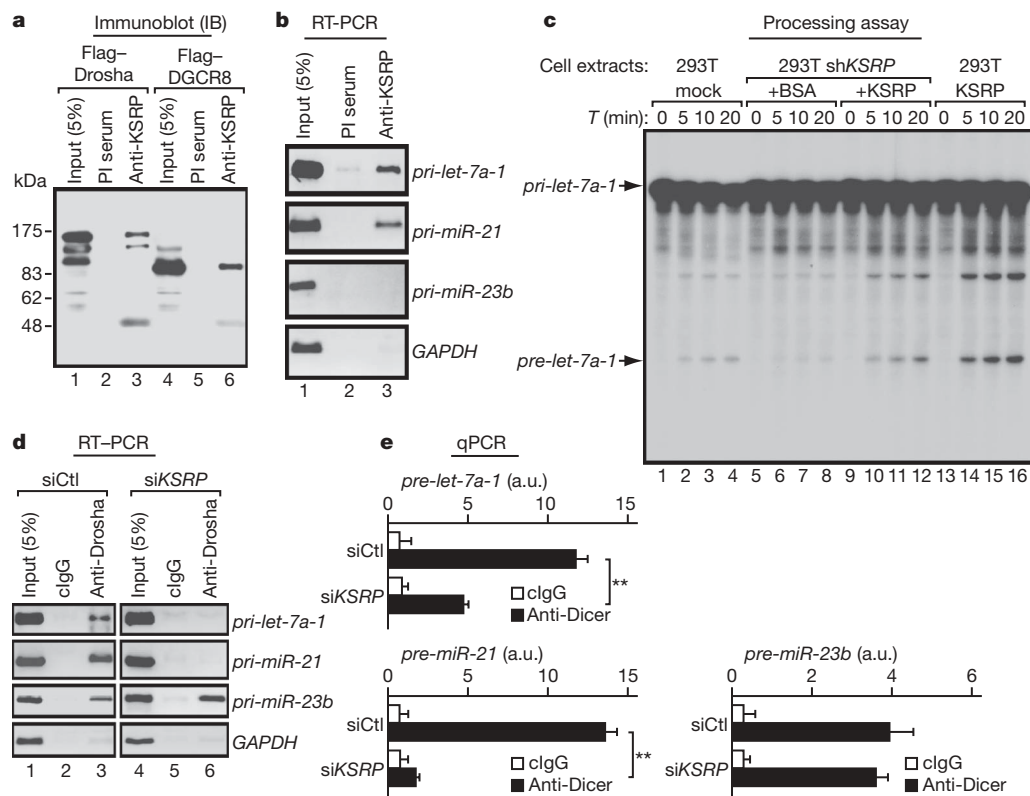

Figure 3 | KSRP is a component of the microprocessor complex, interacts with *pri-let-7a-1* favouring its processing, and is required for interaction of both Drosha and Dicer complexes with let-7a precursors.
a, Coimmunoprecipitation of endogenous KSRP with either Flag–Drosha or Flag–DGCR8. **b**, Anti-KSRP immunoprecipitates select pri-miRNAs. *GAPDH*, glyceraldehyde-3-phosphate dehydrogenase. **c**, KSRP (50 nM)

restores *pri-let-7a-1* processing when added to extracts from sh*KSRP* stably transfected cells whereas KSRP overexpression enhances *pri-let-7a-1* processing. **d**, **e**, *KSRP* knockdown impairs the interaction of Drosha and Dicer with certain miRNA precursors. **$P < 0.01$. All data are presented as mean and s.d. ($n = 4$). cIgG, control immunoglobulin G.
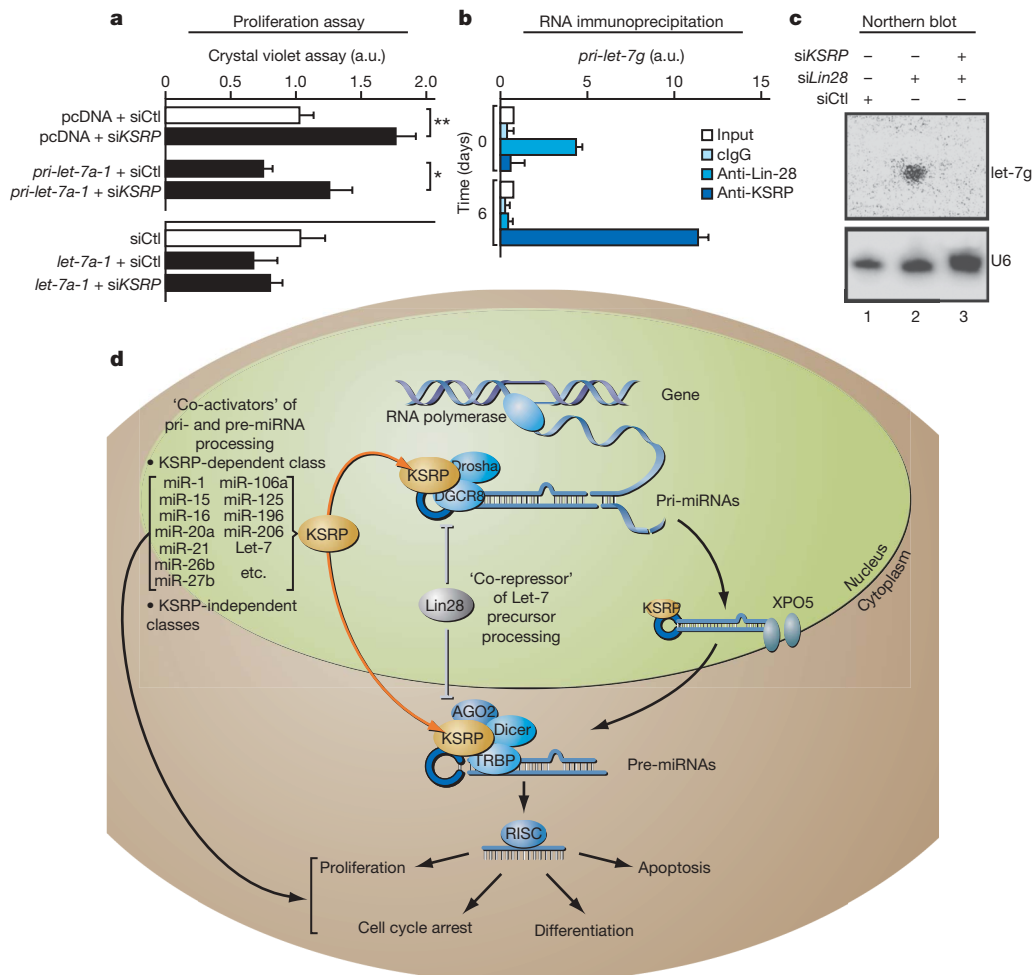
Figure 4 | KSRP affects let-7-regulated cell proliferation and is involved in Lin28-regulated maturation of let-7g in P19 cells. a, *KSRP* knockdown increases proliferation and inhibits the antiproliferative effect of transfected *pri-let-7a-1* but not of mature let-7a in U2OS cells. *P < 0.05, **P < 0.01. **b,** KSRP interacts with *pri-let-7g* only in differentiated P19 cells. **c,** Lin28 knockdown in undifferentiated P19 cells induces the expression of let-7g. Concomitant Lin28 and KSRP knockdown abolishes let-7g expression. **d,** A model for KSRP-dependent regulation of processing of certain miRNAs. All data are presented as mean and s.d. (*n* = 4).

(also known as connexin 43) and DNA pol α[22], impairing C2C12 differentiation (Supplementary Fig. 12e, f).

Recently, four papers[8,23–25] demonstrated that the maturation of let-7 is blocked by Lin28 in undifferentiated embryonic stem cells and P19 cells. We observed that KSRP interacts with *pri-let-7g* in P19 cells upon retinoic-acid-induced differentiation (Fig. 4b). *Lin28* knockdown in undifferentiated P19 cells induced let-7g expression whereas concomitant *KSRP* knockdown abolished this effect (Fig. 4c) indicating that, upon *Lin28* knockdown, KSRP promotes maturation of let-7g precursors. Similarly, upon P19 differentiation, Lin28 expression is abrogated[8,23–25], thus allowing KSRP to promote processing of let-7g precursors. This is also compatible with the recently reported mechanism of action of Lin28 (ref. 26). We suggest that TL is a pivotal structure where miRNA processing 'activators' (for example, KSRP) as well as 'repressors' (for example, Lin28) function in a coordinated way to convey proliferating, apoptotic or differentiating cues into changes of miRNA expression (Fig. 4d).

In conclusion, KSRP is a key regulator of the processing of a sizeable subset of miRNA precursors on the basis of its high-affinity binding to their TL. The TLs of most KSRP-regulated microRNAs (let-7-a, -b, -c, -d, -f, -i and miR-196a) contain short G-rich stretches of at least three Gs that represent the optimal binding site for KH3 (ref. 15). However, the TLs of the other KSRP target miRNAs contain instead two sequential or isolated Gs and our data on the KSRP–TL-miR-21 interaction show that a significant, albeit different, contribution to the binding is provided by both G-containing stretches

(Supplementary Fig. 6d). These data underscore the adaptability of the protein to a broad range of single-stranded RNA sequences[15,16].

Upon binding, KSRP could optimize the positioning and/or recruitment of both the miRNA precursor processing complexes through protein–protein interactions (Supplementary Fig. 13a–c). The RNase sensitivity of KSRP–exportin-5 (XPO5) interaction indicates that KSRP is associated with the TL of target miRNA precursors during nucleo-cytoplasmic transit (Supplementary Fig. 13a). Sequential immunoprecipitation experiments indicate that at least two pools of miRNA precursors exist, one associated with processing complexes including KSRP and the other associated with processing complexes that do not include KSRP (Supplementary Fig. 14). For example, it has been reported[27] that HNRNPA1 binds to the TL of miRNAs for which processing is not affected by KSRP.

Altogether, our findings uncover an additional level of complexity for miRNA-dependent regulation of gene expression that contributes to the modulation of different biological programs.

METHODS SUMMARY

Immunoprecipitation of ribonucleoprotein complexes was performed as described previously[28] with minor modifications. In brief, cells lysates were immunoprecipitated with either protein A- or protein A/protein G–Sepharose-coupled antibodies at 4 °C overnight (16 h). Pellets were sequentially washed with the following buffers: buffer I (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl); buffer II (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl); and buffer III (0.25 M LiCl, 1%

NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1). Total RNA was prepared using Trizol, retrotranscribed using random primers and amplified by PCR. The primer sequences are detailed in Supplementary Table 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Pillai, R. S., Bhattacharyya, S. N. & Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.* **17**, 118–126 (2007).
2. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nature Genet.* **39**, 673–677 (2007).
3. Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
4. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nature Rev. Cancer* **6**, 857–866 (2006).
5. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
6. Obernosterer, G., Leuschner, P. J., Alenius, M. & Martinez, J. Post-transcriptional regulation of microRNA expression. *RNA* **12**, 1161–1167 (2006).
7. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
8. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100 (2008).
9. Gherzi, R. *et al.* A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. *Mol. Cell* **14**, 571–583 (2004).
10. Garcia-Mayoral, M. F. *et al.* The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15**, 485–498 (2007).
11. Haase, A. D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961–967 (2005).
12. Min, H., Turck, C. W., Nikolic, J. M. & Black, D. L. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev.* **11**, 1023–1036 (1997).
13. Kroll, T. T., Zhao, W. M., Jiang, C. & Huber, P. W. A homolog of FBP2/KSRP binds to localized mRNAs in Xenopus oocytes. *Development* **129**, 5609–5619 (2002).
14. Roush, S. & Slack, F. J. The let-7 family of microRNAs. *Trends Cell Biol.* **18**, 505–516 (2008).
15. Garcia-Mayoral, M. F., Diaz-Moreno, I., Hollingworth, D. & Ramos, A. The sequence selectivity of KSRP explains its flexibility in the recognition of the RNA targets. *Nucleic Acids Res.* **36**, 5290–5296 (2008).
16. Ruggiero, T. *et al.* Identification of a set of KSRP target transcripts upregulated by PI3K-AKT signaling. *BMC Mol. Biol.* **8**, 28 (2007).
17. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
18. Sampson, V. B. *et al.* MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res.* **67**, 9762–9770 (2007).
19. Lee, Y. S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* **21**, 1025–1030 (2007).
20. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl Acad. Sci. USA* **102**, 13944–13949 (2005).
21. Chen, J. F. *et al.* The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genet.* **38**, 228–233 (2006).
22. Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A. & Dutta, A. Muscle-specific microRNA miR-206 promotes muscle differentiation. *J. Cell Biol.* **174**, 677–687 (2006).
23. Newman, M. A., Thomson, J. M. & Hammond, S. M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539–1549 (2008).
24. Rybak, A. *et al.* A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nature Cell Biol.* **10**, 987–993 (2008).
25. Piskounova, E. *et al.* Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* **283**, 21310–21314 (2008).
26. Heo, I. *et al.* Lin28 mediates the terminal uridylation of let-7 precursor microRNA. *Mol. Cell* **32**, 276–284 (2008).
27. Michlewski, G., Guil, S., Semple, C. A. & Caceres, J. F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–393 (2008).
28. Chen, C. Y. *et al.* Nucleolin and YB-1 are required for JNK-mediated interleukin-2 mRNA stabilization during T-cell activation. *Genes Dev.* **14**, 1236–1248 (2000).

METHODS

Identification of Dicer-interacting proteins. Anti-Dicer monoclonal antibodies 33, 73 and 83 and control monoclonal antibodies[11] were crosslinked to protein G–Sepharose 4 Fast Flow (Amersham Bioscience) and used to purify Dicer complexes from HEK293T (293T) cytoplasmic extracts (S10, ref. 11). Coimmunoprecipitates were washed five times with lysis buffer (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.5% NP-40, 2.5 mM MgCl$_2$). Proteins were separated by 10% SDS–PAGE. Protein-containing gel fragments were digested with trypsin as described previously[29] and analysed by liquid chromatography tandem mass spectrometry (LCQ Deca XP, Thermo 7 Finnigan). Proteins were identified using Turbo Sequest and MASCOT, searching the SwissProt database restricted to human proteins. A protein was considered as 'identified' if at least two peptides were in the first rank, concerning the correlation of experimental with theoretical data, with an ion score greater than 25. Peptides with ion scores of between 20 and 40 were peer-reviewed for their quality of alignment. No KSRP-specific peptides were identified in immunoprecipitations with isotype-control monoclonal antibodies.

NMR and circular dichroism spectroscopy. All circular dichroism spectra were recorded on a Jasco J-715 spectropolarimeter (Jasco) equipped with a PTC-348 Peltier temperature-control system. RNA binding was monitored by adding increasing amounts of protein to 2 μM *TL-let7a-1* RNA in 10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.5 mM TCEP. A temperature of 5 °C was chosen to optimize the signal change upon protein binding. The integral of the signal of between 255 nm and 265 nm was fitted against the protein concentration using in-house programs described previously[30] and the K_d values were extracted. All NMR spectra were recorded on Bruker Avance spectrometers operating at 600 and 700 MHz ^{1}H frequencies fitted with a supercooled probe. The spectra were processed with the NMRPipe package[31] and analysed with Sparky[32]. Solutions of 25 μM ^{15}N-labelled samples of KH3, KH4 and KH3–4 in 10 mM Tris-HCl buffer, 50 mM NaCl, 1 mM TCEP, pH 7.4, were titrated with *TL-let-7a-1* RNA oligonucleotides. ^{15}N-^{1}H HSQC spectra were recorded at each point of the titration at 27 °C. Amide chemical shift changes as a function of RNA/protein ratio were fitted to obtain the K_d values for the complexes using in-house software as described previously[30]. Weighted average values of ^{15}N and ^{1}H chemical shift variations have been calculated by $\Delta\delta_{av} = [(\Delta\delta^1H)^2 + (\Delta\delta^{15}N/10)^2]^{1/2}$, and used to map *TL-let-7a-1* binding on a MolMol-generated molecular surface[33].

Preparation of RNA substrates and *in vitro* processing assays. For pri-miRNA processing assays, total cell extracts were prepared in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% Triton X-100, 1× Complete, 10% glycerol from either HeLa or 293T cells and incubated (typically 40 μg per 25-μl reaction at 37 °C for the indicated times) with *in vitro* synthesized and uniformly labelled pri-miRNAs (5 fmol) in processing buffer containing 100 mM KCH$_3$COOH, 2 mM Mg(CH$_3$COOH)$_2$, 10 mM Tris-Cl, pH 7.6, 2 mM DTT, 10 mM creatine phosphate, 1 μg creatine phosphokinase, 1 mM ATP, 0.4 mM GTP, 0.1 mM spermine, 2 units of RNasin. Pre-miRNA processing assays were performed as described previously[11].

miRNA profiling. HeLa cells were transiently transfected with either siRNA against KSRP or siRNA against luciferase. Total RNA was prepared using Trizol (Invitrogen) and enriched with RNA smaller than 40 nucleotides using PureLink miRNA isolation kit. RNA was labelled with either Cy-3 or Cy-5 using the NCode miRNA labelling system. A dye-swap design was used. Labelled miRNAs were hybridized, in triplicate, to the Invitrogen NCode MultiSpecies miRNA Microarray V1. Data were collected using GenePixPro 5.0 Agilent Software (Axon Instruments). Normalization and data analysis were performed using a bootstrapping method[34].

Recombinant proteins and antibodies. Recombinant Dicer was purified as described previously[35]. MBP–TRBP was expressed in bacteria and purified as described previously[36]. Production of recombinant KSRP and its deletion mutants as well as p37AUF1 has been described previously[9]. Affinity-purified rabbit polyclonal anti-KSRP antibody, and rabbit polyclonal anti-Dicer (349), were described previously[9,37]. Mouse monoclonal anti-GST was purchased from Chemicon. Mouse monoclonal anti-Flag (M2) and anti-α- and -β-tubulin were from Sigma. Rabbit polyclonal anti-Drosha (07-717) was purchased from Upstate, anti-DGCR8 (N-19) goat polyclonal IgG from Santa Cruz and anti-Lin28 goat polyclonal IgG from R&D Systems.

Plasmids. Human *pre-let-7a-1* and *pre-miR-23b* were cloned into pSUPER-gfp-neo plasmid (Oligoengine). Chimaeric constructs including either the TL sequence of miR-23b in the backbone of *pre-let-7a-1* (UGAGGUAGUAGGUUGUAU AGUUGUGACUUAAGAUUAAACUAUACAAUCUACUGUCUUUC) or the TL sequence of let-7a-1 in the backbone of *pre-miR-23b* (UGGGUUCCUG GCAUGCUGAUUUUUAGGGUCACACCCACCACUGGGAGAUAAAUCACA UUGCCAGGGAUUACC) were generated and cloned into pSUPER-gfp-neo. pCY vector[38] containing part of the sequence (encompassing the mature

miRNA) of either *pri-let-7a-1* or *pri-miR-23b* was generated. Three miR-23b-binding sites from a region between 360 and 385 nucleotides of semaphorin 6D 3′ untranslated region were cloned in pMIR-REPORT plasmid (Ambion). This sequence is a potential target of miR-23b according to TargetScan prediction program (http://www.targetscan.org). Flag–Dicer (provided by M. Doyle) and Flag–TRBP contain a triple Flag tag at the amino terminus of proteins expressed from the pCIneo vector (A.D.H., unpublished). Flag–KSRP was described previously[39]. A luciferase reporter gene plasmid containing six let-7a-binding sites and a plasmid containing genomic sequence encoding *pri-let-7a* were gifs from J. G. Belasco. Flag–Drosha and Flag–DGCR8 were provided by V. Narry Kim. Myc-tagged exportin-5 was provided by Addgene (Addgene plasmid number 12552). Part of the genomic sequence encompassing mouse *pre-miR-1-2* (*pri-miR-1-2*) was cloned into pcDNA3 vector (Invitrogen). Mouse *pre-miR-1-2* was cloned into pSUPER-gfp-neo plasmid (Oligoengine). A luciferase reporter gene plasmid containing four miR-1-binding sites was provided by D. Srivastava.

Cell transfection, coimmunoprecipitation and immunoblotting. HeLa, U2OS, P19 or NIH-3T3 cells were transiently transfected with Lipofectamine 2000 (Invitrogen). After either 48 h or 72 h, cells were collected, washed with PBS, resuspended in lysis buffer, and the protein concentration of cell extracts determined by the Dc Protein assay (Bio-Rad). When required, cell lysates were incubated at room temperature with either RNase A (10 μg ml^{-1}, Ambion) or RNase V1 (1 U ml^{-1}, Ambion) for 30 or 15 min, respectively. Three-hundred micrograms of protein was immunoprecipitated with protein A–Sepharose-bound anti-KSRP antibody for 16 h at 4 °C with rotation. Immunoprecipitates were washed four times with lysis buffer and resuspended in SDS protein loading buffer. Proteins were subjected to SDS–PAGE, electroblotted onto PVDF membranes, and probed with antibodies as indicated. Anti-connexin 43 antibody was from Sigma, anti-DNA pol α (G-16) was from Santa Cruz and anti-myogenin (F5D) was from Iowa Hybridoma Bank. C2C12 myoblasts were cultured and treated as described previously[40] and transfected using Lipofectamine 2000.

Northern blot analysis. RNA was extracted using Trizol, resolved on 15% polyacrylamide–urea gels, and electroblotted onto HyBond N$^+$ membranes. Membranes were hybridized overnight with radiolabelled antisense miRNAs in ExpressHyb solution (Clontech). After hybridization, membranes were washed three times with 2× SSC and 0.05% SDS, twice with 0.1× SSC and 0.1% SDS, exposed overnight to imaging screens, and analysed using a Storm 860 PhosphorImager. Signals were quantified using Imagequant V1.2. The same blot was hybridized (upon stripping in boiling 0.1% SDS) with three distinct probes, including control U6 RNA.

Luciferase assay. Either HeLa or NIH-3T3 cells (80% confluence in 12-well plates) were transfected with Luc reporters containing either six let-7a-binding sites (pGL3-let-7a6XBS) or three miR-23b-binding sites (pGL3-miR-23b3XBS), or empty pGL3 plasmid together with siRNAs using Lipofectamine 2000. For some experiments, cells were co-transfected with either *pri-let-7a-1* or *pre-let-7a-1* expression plasmids or mature let-7a. A co-transfected β-galactosidase-containing plasmid was used to normalize firefly luciferase activity.

EdU incorporation for proliferation assay. Seventy-two hours after transfection, cells were incubated for 2 h with EdU-containing medium. Nuclear incorporation of EdU was determined using Click-iT EdU imaging kit (Invitrogen).

Crystal violet proliferation assay. Seventy-two hours after transfection, U2OS cells were fixed and stained with crystal violet solution. After two washes with water, crystal violet staining was measured by spectrophotometer at a wavelength of 590 nm.

Tunel assay. Tunel-fluorescent staining of apoptotic cells was performed on U-2OS 72 h after the transfection according to the manufacturer's protocol (Roche).

Ultraviolet-crosslinking experiments and gel mobility shift assays. Ultraviolet-crosslinking experiments and gel mobility shift assays were performed essentially as described[9,38].

Expression and purification of the recombinant proteins used for biophysical studies. KSRP KH3, KH4 and KH3–4 proteins were obtained as described previously[10]. In brief, ^{15}N-labelled proteins were expressed in *Escherichia coli* BL21 (DE3), as His–GST fusion protein, and initially purified using nickel affinity chromatography according to the manufacturer's instructions. The His–GST fusion tags were then cleaved with TEV protease and removed using a second nickel affinity step. Proteins were further purified and buffer exchanged by gel filtration (Superdex 75 16/60 column, Pharmacia). Protein purity (always >95%) was assessed using SDS–PAGE and Coomassie blue staining. Protein quantification was achieved by a combination of spectrophotometry using predicted extinction coefficients and ninhydrin analysis of protein hydrolysates. All RNA oligonucleotides were chemically synthesized (Dharmacon).

shRNA-mediated KSRP knockdown. To knockdown stably KSRP, the following oligonucleatides 5′-GATCAACCGGAGAGCAAGA-3′ and 5′-GGACAG TTTCACGACAACG-3′ for human and mouse proteins, respectively, were

cloned into pSUPER-Puro (Oligoengine). 293T, NIH-3T3 or C2C12 cells were transfected using Lipofectamine Plus (Invitrogen). Transfectant pools were kept under selection in medium containing either $3\,\mu g\,ml^{-1}$ (293T) or $1.2\,\mu g\,ml^{-1}$ puromycin (NIH-3T3).

RT–PCR and quantitative PCR. PCR with reverse transcription (RT–PCR) and quantitative PCR were performed as described previously[41].

siRNAs. To knockdown the following human proteins, these siRNAs were synthesized by Qiagen: human KSRP 5′-GAUCAACCGGAGAGCAAGAUU-3′, mouse KSRP 5′-GGACAGUUUCACGACAACG-3′, mouse Lin28 5′-GGGUU GUGAUGACAGGCAA-3′, human Dicer 5′-GAAUCAGCCUCGCAACAAA UU-3′, luciferase 5′-CGUACGCGGAAUACUUCGAUU-3′.

29. Schrimpf, S. P., Langen, H., Gomes, A. V. & Wahlestedt, C. A two-dimensional protein map of Caenorhabditis elegans. *Electrophoresis* **22**, 1224–1232 (2001).
30. Martin, S. R. *et al.* Interaction of calmodulin with the phosphofructokinase target sequence. *FEBS Lett.* **577**, 284–288 (2004).
31. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
32. Goddard, T. D. & Kneller, D. G. *SPARKY, University of California, San Francisco* ⟨http://www.cgl.ucsf.edu/home/sparky/⟩.
33. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **51-5**, 29–32 (1996).
34. Kerr, M. K. & Churchill, G. A. Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201 (2001).
35. Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
36. Dorin, D. *et al.* The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs in vitro and in vivo independently of its ability to inhibit the dsRNA-dependent kinase PKR. *J. Biol. Chem.* **278**, 4440–4448 (2003).
37. Kotaja, N. *et al.* The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc. Natl Acad. Sci. USA* **103**, 2647–2652 (2006).
38. Chen, C. Y. *et al.* AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**, 451–464 (2001).
39. Briata, P. *et al.* p38-dependent phosphorylation of the mRNA decay-promoting factor KSRP controls the stability of select myogenic transcripts. *Mol. Cell* **20**, 891–903 (2005).
40. Wu, Z. *et al.* p38 and extracellular signal-regulated kinases regulate the myogenic program at multiple steps. *Mol. Cell. Biol.* **20**, 3951–3964 (2000).
41. Gherzi, R. *et al.* The RNA-binding protein KSRP promotes decay of beta-catenin mRNA and is inactivated by PI3K-AKT signaling. *PLoS Biol.* **5**, e5 (2006).

nature

LETTERS

Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase

Thomas R. M. Barends[1], Elisabeth Hartmann[1], Julia J. Griese[1], Thorsten Beitlich[1], Natalia V. Kirienko[2], Dmitri A. Ryjenkov[2], Jochen Reinstein[1], Robert L. Shoeman[1], Mark Gomelsky[2] & Ilme Schlichting[1]

The ability to respond to light is crucial for most organisms. BLUF is a recently identified photoreceptor protein domain that senses blue light using a FAD chromophore[1]. BLUF domains are present in various proteins from the Bacteria, Euglenozoa and Fungi. Although structures of single-domain BLUF proteins have been determined[2–4], none are available for a BLUF protein containing a functional output domain; the mechanism of light activation in this new class of photoreceptors has thus remained poorly understood. Here we report the biochemical, structural and mechanistic characterization of a full-length, active photoreceptor, BlrP1 (also known as KPN_01598), from *Klebsiella pneumoniae*[5]. BlrP1 consists of a BLUF sensor domain and a phosphodiesterase EAL output domain which hydrolyses cyclic dimeric GMP (c-di-GMP). This ubiquitous second messenger controls motility, biofilm formation, virulence and antibiotic resistance in the Bacteria[6–9]. Crystal structures of BlrP1 complexed with its substrate and metal ions involved in catalysis or in enzyme inhibition provide a detailed understanding of the mechanism of the EAL-domain c-di-GMP phosphodiesterases. These structures also sketch out a path of light activation of the phosphodiesterase output activity. Photon absorption by the BLUF domain of one subunit of the antiparallel BlrP1 homodimer activates the EAL domain of the second subunit through allosteric communication transmitted through conserved domain–domain interfaces.

BlrP1 shows a typical BLUF-domain photocycle[5] and c-di-GMP hydrolytic activity (Table 1), making it a good candidate for investigating signal transduction in BLUF-domain proteins and understanding the mechanism of EAL-domain proteins. We crystallized BlrP1 as the calcium-inhibited c-di-GMP complex in the dark; other metal complexes were obtained from soaking experiments. The BlrP1 structure

was determined from selenium single-wavelength anomalous dispersion data to 2.4 Å resolution (Supplementary Table 1). The asymmetric unit of the crystals contains two antiparallel BlrP1 molecules (Fig. 1). At the core of the BlrP1 dimer, the two EAL domains interact through three helices: one from each EAL domain, which run antiparallel; and one 'compound helix' consisting of two helical segments, one from each EAL domain, whose amino termini meet end to end. The dimer interface (4,900 Å²) is made up of conserved residues (Supplementary Fig. 1), suggesting its functional relevance. In fact, the same kind of dimer is found in other EAL-domain-containing structures, *Bacillus subtilis* YkuI protein (Protein Data Bank accession code 2BAS)[10] and *Thiobacillus denitrificans* TBD1265 (Protein Data Bank accession code 2R6O)[11], which were determined in the absence of substrate. Size-exclusion chromatography, dynamic light scattering and analytical

Table 1 | Rates of hydrolysis of c-di-GMP to linear diguanylate, 5′pGpG, for YcgF and BlrP1

Protein	Conditions	k_{cat} (s^{-1}), dark conditions	k_{cat} (s^{-1}), light conditions
Wild-type YcgF	Standard*	Below detection	Below detection
Wild-type BlrP1	Standard	0.13 ± 0.02	0.54 ± 0.02
Wild-type BlrP1	pH 6.0	Below detection	Below detection
Wild-type BlrP1	pH 8.0	1.84 ± 0.16	2.54 ± 0.10
Wild-type BlrP1	pH 9.3	6.70 ± 1.80	6.17 ± 0.92
Wild-type BlrP1	10 mM MnCl₂	0.64 ± 0.16	1.17 ± 0.19
Wild-type BlrP1	10 mM CaCl₂	Below detection	Below detection
Wild-type BlrP1	10 mM ZnCl₂	Below detection	Below detection
Wild-type BlrP1	10 mM CuCl₂	Below detection	Below detection
BlrP1 R93S	Standard	0.27 ± 0.01	1.42 ± 0.08
BlrP1 L128C–G353C	Standard	2.06 ± 0.24	3.48 ± 0.28
BlrP1 S309C–S312C	Standard	Below detection	Below detection

The indicated errors are standard deviations; k_{cat}, catalytic-turnover rate constant.
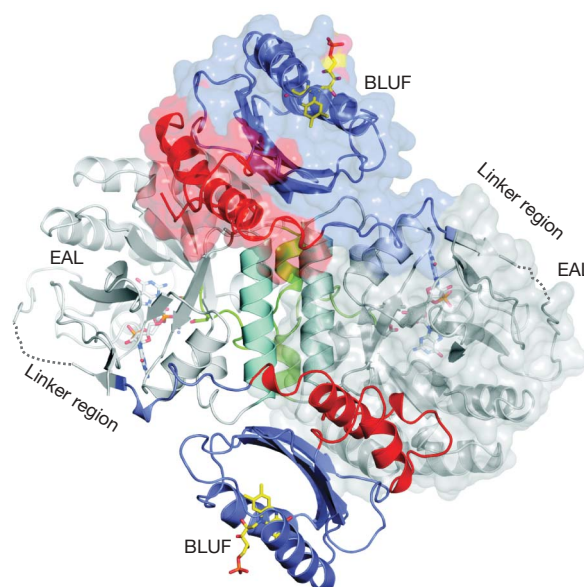* 1.1 µM protein, 100 µM c-di-GMP, 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl₂, 20 °C.

Figure 1 | Overview of a BlrP1 dimer. EAL domains are shown in grey, BLUF domains in blue with a yellow flavin mononucleotide (FMN) molecule. A transparent surface representation is shown for one of the monomers to distinguish the monomers from each other. The carboxy-terminal helices of the BLUF domains (red) contact the EAL domains of the other monomer with α4$_{BLUF}$ (residues 124–134). The conserved helices α6 of the EAL domains are involved in dimerization and shown in cyan. The two helices making up the compound helix in the dimer interface (see text) are shown in green, as is the loop connecting these helices to the EAL active sites. The FMN chromophores and the c-di-GMP substrates have similar intra- and intermonomer distances (42 Å). The interface between the BLUF and EAL domains of different monomers suggests an intermolecular activation mechanism.

[1]Max Planck Institute for Medical Research, Department of Biomolecular Mechanisms, Jahnstrasse 29, 69120 Heidelberg, Germany. [2]Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 82071, USA.
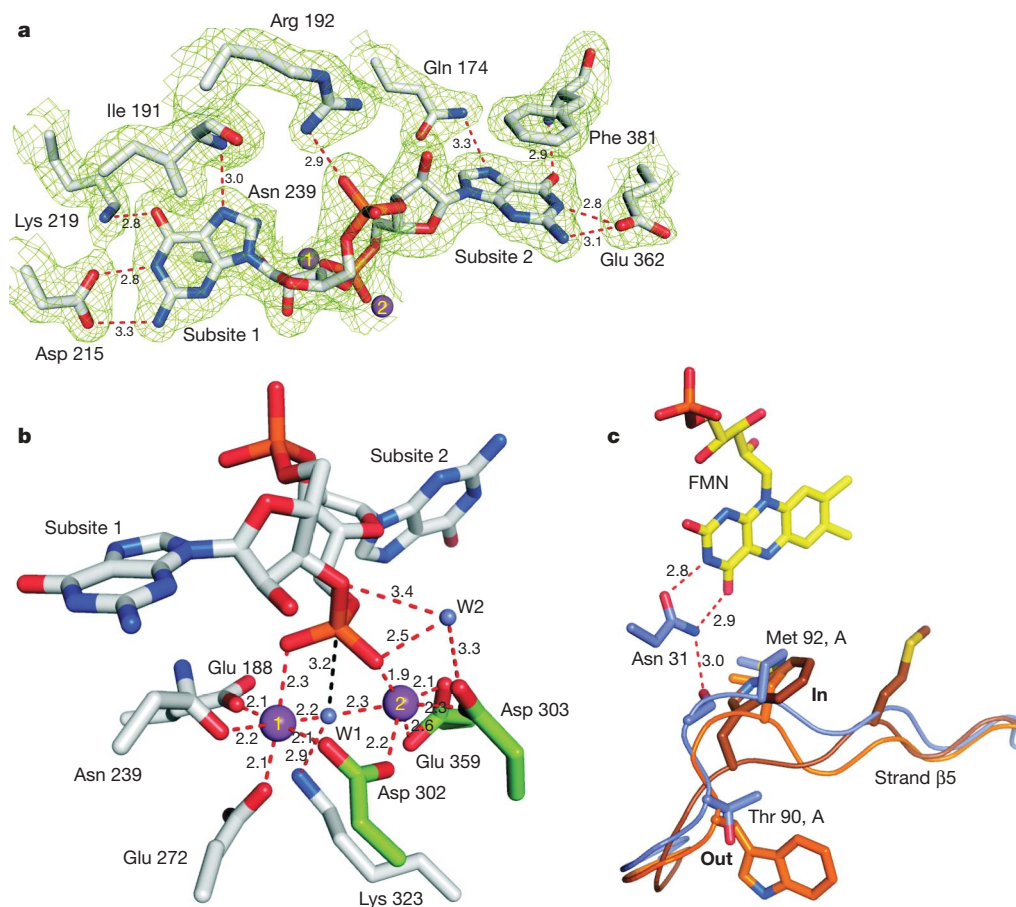
1015

Figure 2 | Detailed views of the active sites of BlrP1. a, The c-di-GMP binding site of BlrP1 in the B molecule of the manganese complex at pH 6. Distances are indicated in angstroms. The $2mF_o - DF_c$ electron density is shown in green, contoured at 1.0σ. **b**, Enlarged view of the EAL active site of the B monomer of BlrP1 determined from a crystal soaked with Mn^{2+} ions at pH 9. The manganese ions (violet) are coordinated by several carboxylic acid residues including Glu 359, located on strand β7$_{EAL}$, and Asp 302 and Asp 303, located on a loop leading to the dimer interface (green; see text). Also shown are the putative hydrolytic solvent molecule, W1, in a possible near-attack conformation[28] and a possible leaving-group proton donor, water molecule W2. **c**, Superposition of three BLUF domain structures: BlrP1 (blue), AppA in the Met$_{in}$ state (light brown)[21] and AppA in the Met$_{out}$ state (dark brown)[22]. BlrP1 and *E. coli* YcgF do not have Trp two residues before the conserved Met, but instead have Thr and Ala, respectively. BlrP1 clearly adopts the Met$_{in}$ state, with Met 92 located in the binding pocket and Thr 90 pointing towards the solvent (monomer A). Strand β5 is not extended but instead has a kink that is stabilized by a hydrogen bond between the backbone of strand β5 and Asn 31, which binds to the flavin (yellow, modelled as FMN).

ultracentrifugation demonstrate that BlrP1 and its *Escherichia coli* homologue, YcgF (ref. 12), are also dimeric in solution (Supplementary Figs 2 and 3).

The central β-sheet of the N-terminal BlrP1 BLUF domain (residues 1–98) is capped off by two C-terminal helices (99–141) that are rotated by 90° relative to single-domain BLUF proteins[2–4] (Supplementary Fig. 4 and Supplementary Table 2)[13]. A linker 50 Å long connects to the C-terminal EAL domain (169–405), which has a triosephosphate isomerase (TIM)-barrel fold (Supplementary Fig. 6). Cyclic di-GMP binds at the top of the barrel in an extended conformation (Fig. 2a), in contrast to the U-shaped conformation observed in the c-di-GMP receptor PilZ domain[14]. As expected for a phosphodiesterase, BlrP1 binds c-di-GMP through two metal ions, Me1 and Me2 (Fig. 2b). In the Mn^{2+} complex at pH 9, which corresponds to the enzymatically highly active conformation of the BlrP1 phosphodiesterase (Table 1), Me1 is coordinated octahedrally by Glu 188 (of the conserved eponymous EAL motif); Asn 239; Glu 272; one phosphate oxygen atom of c-di-GMP; a bridging solvent molecule, W1; and the metal-bridging ligand Asp 302. Me2 is coordinated in a distorted trigonal–bipyramidal fashion by the bridging solvent molecule, the bridging ligand Asp 302, another phosphate oxygen of c-di-GMP, Asp 303 and Glu 359. This configuration suggests that EAL proteins use the same reaction mechanism as postulated for other cyclic nucleotide phosphodiesterases[15,16], with the bridging solvent molecule

performing nucleophilic attack on the phosphorus. In BlrP1, this leads to expulsion of O3 after its protonation by a water molecule and product formation. A potential proton-donating water molecule, W2, is bound to O3 and Asp 303 in one monomer of the Mn^{2+} complex at pH 9.
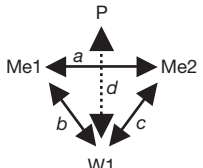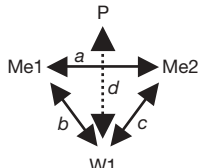
BlrP1 possesses c-di-GMP-phosphodiesterase activity, in contrast to *E. coli* YcgF (Table 1), in which the Me2- and W1-coordinating Glu 359 is replaced with Met 362 (Supplementary Table 3), and Ykul, in which the equivalent of the Me2-coordinating Asp 303 is replaced with Asn[10]. The phosphodiesterase activity of BlrP1 and other EAL proteins increases with pH[17,18] and is abolished in the presence of calcium[17–19] (Table 1). To investigate the structural cause, we determined crystal structures of BlrP1, at different pH values, complexed with calcium or manganese ions. Table 2 shows the distance between the two active-site metal ions, and their distances to the bridging solvent molecule, W1, positioned for in-line attack of the phosphorus atom. In the highly active (Table 1) high-pH Mn^{2+} complex, a tight binuclear metal centre is formed in which both metals coordinate W1, which is presumably transformed to a hydroxide in the catalytically active complex[15,20]. In the inactive calcium and low-pH complexes (Table 1), the metal ions are further apart, and only Me1 coordinates W1, which is insufficient for its activation. For maximal phosphodiesterase-rate acceleration, both the active-site geometry and the occupancy of the hydroxide need to be optimized[20]. The importance of deprotonation of the metal-ligating

water to afford hydroxide is supported by the higher enzymatic activity with Mn^{2+} than with Mg^{2+}, which correlates with higher Lewis acidity of Mn^{2+}. The second feature most likely contributing to the increase in activity with pH is the protonation state of the ε-amino group of Lys 323, orienting and activating solvent molecule W1.

Comparing the high-pH Mn^{2+} complex with the low-pH and Ca^{2+} complexes, major structural differences are found in helix $\alpha4_{EAL}$ (275–278), notably at Glu 275, which interacts with either Gly 305 or Gly 310, which are respectively located before and after the turn before the compound helix ($\alpha5_{EAL}$, 311–315). This structural feature is mirrored by the adjacent helix $\alpha4_{EAL}$ and compound helix $\alpha5_{EAL}$ of the other monomer. The differences in helix $\alpha4_{EAL}$ affect the $\beta5_{EAL}$–$\alpha5_{EAL}$ loop containing Asp 302 and Asp 303, and thus the Me2 coordination and the position and coordination of W1 and W2. In the active conformation, Me2 moves towards Glu 359, whose side-chain carboxylate rotates by ~90° to coordinate Me2 and W1 instead of interacting with the backbone amide of Asp 303. The importance of these regions for phosphodiesterase activity is indicated by the observations that the S309C–S312C double mutant shows no activity (Table 1) and that in the EAL phosphodiesterase RocR from *Pseudomonas aeruginosa*, substitution of Ala and Gln for the residue equivalent to BlrP1's Glu 275 resulted in >10^5-fold and 446-fold decreases in enzymatic activity, respectively[18] (Supplementary Table 3).

Light increases the phosphodiesterase activity of BlrP1 fourfold (Table 1), suggesting that light-induced structural changes originating in the BLUF domain are transmitted to the EAL domain to optimize its active site for catalysis. We propose that light activation increases the overall occupancy of a state with a properly coordinated Me2 involving the structural changes outlined above. The critical issue is that of how light-induced changes in BLUF are transferred to the EAL domain. Clues come from the observation of two conformations of BLUF domain (Met_{in}, Met_{out}; ref. 21) that differ in the position of key residues located in the flavin-binding pocket[2,4,13,21–23] (Fig. 2c and Supplementary Fig. 5) and that have been correlated with the light-induced changes. In what we proposed to be the dark-adapted state of the BLUF photoreceptor[21,24], the strictly conserved Met residue (here Met 92) is located in the flavin-binding pocket next to the conserved Gln (here Gln 49), whereas Thr 90 points towards the solvent (Met_{in} conformation). Many BLUF proteins contain Trp in place of Thr 90, whence the earlier 'Trp_{out}, Trp_{in}'[21] terminology, which is less universal than that used here. Absorption of a photon by the flavin is believed to lead to a rotation[22,25] and tautomerization[24,26] of Gln that alters its interactions with the Met residue. Met then flips out of the flavin-binding pocket and Thr (or Trp in other proteins) takes its place (Met_{out} conformation[21]); this in turn straightens the kink in strand β5. The sliding and extension of strand β5 would affect the loop connecting strands β4 and β5 and the orientation of the C-terminal helices, consistent with recent solution studies on the BLUF protein BlrB[23] and the BlrP1 BLUF domain[13].

Table 2 | Distances between Me1, Me2, W1 and the attacked phosphorous atom P

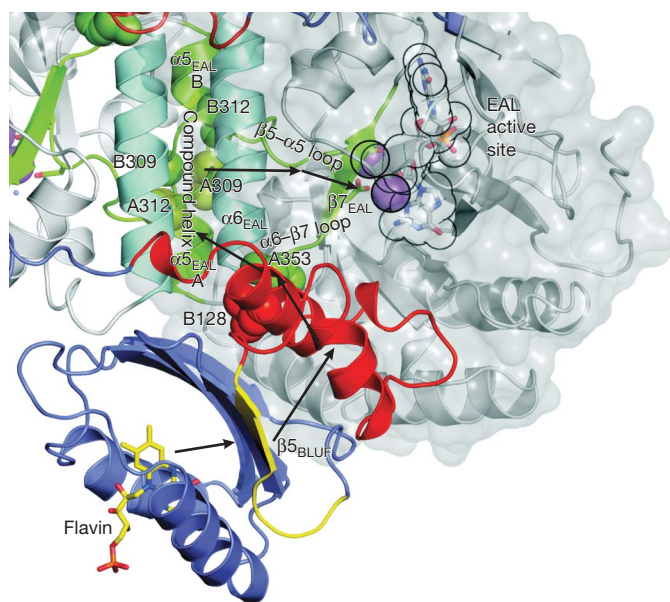| Complex	Molecule A (crystal contact)	Molecule B (open active site)
	a, b, c, d (Å)	*a, b, c, d* (Å)
Mn^{2+}, pH 6.0	4.2, 2.3, 3.0, 3.4	4.3, 2.2, 3.4, 3.4
Mn^{2+}, pH 9.0	3.8, 2.4, 2.4, 3.5	3.7, 2.3, 2.3, 3.2
Ca^{2+}, pH 4.5	—, 2.6, —, 3.7	—, 2.4, —, 3.9
Ca^{2+}, pH 8.0	4.9, 2.4, 4.1, 3.6	5.9, 2.4, 4.7, 3.4

Figure 3 | Proposed light-regulation pathway between the BLUF and EAL domains of BlrP1 (arrows). Light absorption by the flavin (yellow) results in a rearrangement of strand $\beta5_{BLUF}$ and the $\beta4_{BLUF}$–$\beta5_{BLUF}$ loop (both yellow), as described in the text. This affects the C-terminal helices of the BLUF domain (red, monomer B), of which $\alpha4_{BLUF}$ contacts the EAL domain of the other monomer (monomer A) at the α6–β7 loop (green) between the dimerization helix $\alpha6_{EAL}$ (cyan) and EAL strand $\beta7_{EAL}$ (green). From here the signal reaches the dimerization helix $\alpha6_{EAL}$ and the compound helix $\alpha5_{EAL}$ (green). The elements of the compound helix are attached to the β5–α5 loop (green) containing Asp 302 and Asp 303 (sticks), which coordinate the metal ions (purple spheres) of the active site. The metal ions and c-di-GMP substrate (sticks) are additionally indicated with a black outline. Mutation sites (Table 1) are shown as spheres, and residues from the two monomers are indicated by A and B, as appropriate. See also Supplementary Fig. 7.

These proposed light-induced changes affect the BLUF–EAL interface in BlrP1, where the C-terminal helix of the BLUF domain of one monomer interacts with the connector (351–354) between the EAL–EAL interface forming helix $\alpha6_{EAL}$ and the following β-strand ($\beta7_{EAL}$, 355–359) of the other monomer (Fig. 3 and Supplementary Fig. 7), which contains the active-site Glu 359. The importance of this interaction for EAL regulation is evidenced by the double mutant L128C–G353C (the first in the C-terminal helix of the BLUF domain, the second in the $\alpha6_{EA}$–$\beta7_{EAL}$ connector), which displays high EAL activity under both dark and light conditions (Table 1). Because the BLUF domain of one monomer contacts the EAL domain of the other monomer, the light-induced change of the BLUF–EAL interactions would further affect the quaternary structure, changing the EAL–EAL interface at the dimerization helix as well as the compound helix $\alpha5_{EAL}$ and the loop (302–310) connecting it to $\beta5_{EAL}$, which are implicated in EAL activation by Mn^{2+} and high pH (see above).

This model of the sensor-induced quaternary-structure-mediated EAL activation is further supported by the comparison of the BlrP1 structure with that of YkuI (ref. 10). Whereas BlrP1 has a common N-terminal sensor/C-terminal effector domain organization, YkuI has an N-terminal EAL domain followed by a PAS-like domain[10], which may function as a sensor. Despite this difference, if single EAL domains of the respective dimeric structures are superimposed, not only do the EAL domain dimers superimpose perfectly, but also the PAS-like domains of YkuI end up in the same region as the C-terminal helices of the BLUF domains in BlrP1 (Supplementary Fig. 7). In fact, the areas of contact between the PAS-like and BLUF domains and their respective EAL domains are similar, also involving the connector between the EAL–EAL interface forming helix $\alpha6_{EAL}$ and the following β-strand, $\beta7_{EAL}$. This is consistent with the

intermolecular switching mechanism described above, in which the sensory domain affects the EAL dimer (Supplementary Fig. 8). The modular design of EAL-domain-containing proteins[27] seems rooted in the close spatial proximity of the N and C termini of the polypeptide chains of TIM-barrel domains allowing attachment of other domains at either end while maintaining the same relative positions and, thus, the same activation mechanism.

METHODS SUMMARY

We cloned, expressed and purified BlrP1 essentially as described previously[5], and initially obtained *E. coli* YcgF as a glutathione-*S*-transferase–YcgF fusion[12] and purified it using glutathione affinity- and size-exclusion chromatography. Proteins were reconstituted with FAD and BlrP1 was crystallized at 4 °C in the dark using polyethyleneglycol in the presence of c-di-GMP and Ca^{2+}. We obtained other metal complexes by soaking after removal of Ca^{2+} with EDTA/EGTA buffer. The structure of BlrP1 was solved for using single-wavelength anomalous dispersion data from a selenomethionine-labelled crystal. Oligomeric states were determined using analytical ultracentrifugation, size-exclusion chromatography and dynamic light scattering. Cyclic di-GMP phosphodiesterase activities were measured essentially as described previously[17].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 December 2008; accepted 5 March 2009.

1. Gomelsky, M. & Klug, G. BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem. Sci.* **27**, 497–500 (2002).
2. Jung, A. *et al.* Structure of a bacterial BLUF photoreceptor: insights into blue light-mediated signal transduction. *Proc. Natl Acad. Sci. USA* **102**, 12350–12355 (2005).
3. Kita, A., Okajima, K., Morimoto, Y., Ikeuchi, M. & Miki, K. Structure of a cyanobacterial BLUF protein, Tll0078, containing a novel FAD-binding blue light sensor domain. *J. Mol. Biol.* **349**, 1–9 (2005).
4. Yuan, H. *et al.* Crystal structures of the *Synechocystis* photoreceptor Slr1694 reveal distinct structural states related to signaling. *Biochemistry* **45**, 12687–12694 (2006).
5. Tyagi, A. *et al.* Photodynamics of the blue-light-regulated phosphodiesterase BlrP1 from *Klebsiella pneumoniae* and its photoreceptor BLUF domain. *Chem. Phys.* **354**, 130–141 (2008).
6. Hisert, K. B. *et al.* A glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. *Mol. Microbiol.* **56**, 1234–1245 (2005).
7. Hoffman, L. R. *et al.* Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* **436**, 1171–1175 (2005).
8. Kulasakara, H. *et al.* Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. *Proc. Natl Acad. Sci. USA* **103**, 2839–2844 (2006).
9. Claret, L. *et al.* The flagellar sigma factor FliA regulates adhesion and invasion of Crohn disease-associated *Escherichia coli* via a cyclic dimeric GMP-dependent pathway. *J. Biol. Chem.* **282**, 33275–33283 (2007).
10. Minasov, B. *et al.* Crystal structures of Ykul and its complex with second messenger c-di-GMP suggests catalytic mechanism of phosphodiester bond cleavage by EAL domains. *J. Biol. Chem.* doi:10.1074/jbc.M808221200 (in the press).
11. Chang, C. *et al.* Crystal structure of a putative diguanylate cyclase/phosphodiesterase from *Thiobacillus denitrificans. RCSB PDB: Structure Explorer* ⟨http://www.rcsb.org/pdb/explore/explore.do?structureId=2R6O⟩ (2007).
12. Gomelsky, M. & Kaplan, S. AppA, a redox regulator of photosystem formation in *Rhodobacter sphaeroides* 2.4.1, is a flavoprotein. Identification of a novel FAD binding domain. *J. Biol. Chem.* **273**, 35319–35325 (1998).
13. Wu, Q. & Gardner, K. H. Structure and insight into blue light-induced changes in the BlrP1 BLUF domain. *Biochemistry* **48**, 2620–2629 (2009).
14. Benach, J. *et al.* The structural basis of cyclic diguanylate signal transduction by PilZ domains. *EMBO J.* **26**, 5153–5166 (2007).
15. Zhan, C. G. & Zheng, F. First computational evidence for a catalytic bridging hydroxide ion in a phosphodiesterase active site. *J. Am. Chem. Soc.* **123**, 2835–2838 (2001).
16. Salter, E. A. & Wierzbicki, A. The mechanism of cyclic nucleotide hydrolysis in the phosphodiesterase catalytic site. *J. Phys. Chem. B* **111**, 4547–4552 (2007).
17. Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J. Bacteriol.* **187**, 4774–4781 (2005).
18. Rao, F., Yang, Y., Qi, Y. & Liang, Z.-X. Catalytic mechanism of c-di-GMP specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa. J. Bacteriol.* **190**, 3622–3631 (2008).
19. Ross, P. *et al.* The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinum.* Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J. Biol. Chem.* **265**, 18933–18943 (1990).
20. Herschlag, D. & Jencks, W. P. Catalysis of the hydrolysis of phosphorylated pyridines by Mg(OH)+: a possible model for enzymatic phosphoryl transfer. *Biochemistry* **29**, 5172–5179 (1990).
21. Jung, A., Reinstein, J., Domratcheva, T., Shoeman, R. L. & Schlichting, I. Crystal structures of the AppA BLUF domain photoreceptor provide insights into blue light-mediated signal transduction. *J. Mol. Biol.* **362**, 717–732 (2006).
22. Anderson, S. *et al.* Structure of a novel photoreceptor, the BLUF domain of AppA from *Rhodobacter sphaeroides. Biochemistry* **44**, 7998–8005 (2005).
23. Wu, Q., Ko, W. H. & Gardner, K. H. Structural requirements for key residues and auxiliary portions of a BLUF domain. *Biochemistry* **47**, 10271–10280 (2008).
24. Domratcheva, T., Grigorenko, B. L., Schlichting, I. & Nemukhin, A. V. Molecular models predict light-induced glutamine tautomerization in BLUF photoreceptors. *Biophys. J.* **94**, 3872–3879 (2008).
25. Gauden, M. *et al.* Hydrogen-bond switching through a radical pair mechanism in a flavin-binding photoreceptor. *Proc. Natl Acad. Sci. USA* **103**, 10895–10900 (2006).
26. Sadeghian, K., Bocola, M. & Schütz, M. A conclusive mechanism of the photoinduced reaction cascade in blue light using flavin photoreceptors. *J. Am. Chem. Soc.* **130**, 12501–12513 (2008).
27. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
28. Hur, S. & Bruice, T. C. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J. Am. Chem. Soc.* **125**, 1472–1473 (2003).

Author Contributions M.G. and I.S. designed the project; N.V.K. and M.G. identified and cloned *blrP1*; N.V.K. and D.A.R. purified BlrP1 and performed the initial biochemical analysis; J.J.G. constructed the expression clone and established the purification protocol of the protein used for crystallization and biochemical analysis; E.H. purified, crystallized and characterized wild-type and mutant BlrP1 by enzymatic and dynamic light scattering analysis; T.B. performed the spectroscopic analysis; T.R.M.B. determined the crystal structures; J.R. designed the kinetic experiments; R.L.S. performed the mass spectrometry and analytical ultracentrifugation; and I.S. did the practical crystallographic work. T.R.M.B., M.G. and I.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The atomic coordinates and structure factor amplitudes of the following structures have been deposited in the Protein Data Bank: Ca^{2+} complex of BlrP1 at pH 4.5 (accession code 3GFX), metal-free complex of BlrP1 at pH 6.0 (3GFY), Mn^{2+} complex of BlrP1 at pH 6.0 (3GFZ), Mn^{2+} complex of BlrP1 at pH 9.0 (3GG0) and Ca^{2+} complex of BlrP1 at pH 8.0 (3GG1). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.G. (gomelsky@uwyo.edu) or I.S. (ilme.schlichting@mpimf-heidelberg.mpg.de).

METHODS

Cloning and expression of BlrP1. The *blrP1* (NCBI UID 5342100) gene (encompassing the sequence corresponding to amino acid residues 1–405 of BlrP1) was PCR-amplified using genomic DNA of *Klebsiella pneumoniae* subsp. *pneumoniae* (Schroeter) Trevisan (ATCC 700721D) and the following primers: BlrP1_XbaI_Fwd, 5′-GCTCTAGAATGCTAACCACCCTGATTTACC GC-3′; BlrP1_Hind_Rev, 5′-CCCAAGCTTTTACAGGTCCATTGCCTGGC GC-3′. The PCR product was digested with XbaI and HindIII and cloned into vector pMAL-c2x (New England Biolabs) digested with the same enzymes to generate pMAL-c2x::blrP1, which expresses an MBP (maltose binding protein)–BlrP1 translational fusion. The NdeI–HindIII fragment containing the MBP–BlrP1 fusion from pMAL-c2x::blrP1 was subsequently cloned into pET28a digested with NdeI and HindIII to generate a plasmid that expresses the His6–MBP–BlrP1 fusion. The resulting vector was transformed into *E. coli* XL-10 Gold, and its sequence was verified (Medigenomix). The vector was then transformed into *E. coli* Rosetta 2 (DE3) to express a His6–MBP–BlrP1 fusion protein that could be purified by Ni-NTA affinity chromatography. This fusion protein contains a factor-Xa cleavage site between the MBP and the BlrP1 parts of the fusion protein. Cleavage with factor Xa results in BlrP1 plus eight additional vector-derived amino acid residues at the N terminus.

For expression, 2 l of LB medium containing 30 μg ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol in a 5-l shaking flask were inoculated with 20 ml overnight culture of *E. coli* Rosetta 2 (DE3)/pET-28a(+)::MBP–BlrP1, and 3–4 drops of antifoaming agent (Sigma) were added. The culture was incubated at 37 °C and 110 r.p.m. until the absorbance at 600 nm, A_{600}, reached ~0.5. Then the incubator was slowly cooled to 18 °C. At an A_{600} of ~1, IPTG was added to a final concentration of 0.5 mM. Expression was continued for 24 h at 18 °C and 110 r.p.m. Cells were collected by centrifugation at 4 °C, 14,000g_{max} (g_{max}, *g* force at maximum radius), in a Sorvall SLC-6000 rotor for 10 min. Two litres of culture yielded a yellowish cell pellet of 7.5–10 ml volume. The cell pellets were frozen at −20 °C. For the production of SeMet-labelled protein, 5 l of minimal medium supplemented with 50 mg l^{-1} l-selenomethionine (Sigma) and the appropriate antibiotics were used.

Purification of BlrP1. Two tablets of protease inhibitor cocktail (Complete, EDTA-free, Roche) were dissolved in ~300 ml of 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5 mM MgCl$_2$, 20 mM imidazole and 5% glycerol to yield lysis buffer. A cell pellet from ~2 l of culture was thoroughly resuspended in 30 ml of this buffer, and the cells were lysed by three passes through the microfluidizer at a pressure of 0.8 MPa. Thereafter, the cell lysate was kept in the dark, and all manipulation, including crystallization and crystal mounting, were conducted in the dark or under red light (unless the yellow colour of the protein had to be seen, for example for buffer exchange) at 4 °C.

Cell debris was pelleted by ultracentrifugation at 4 °C, 186,000g_{max}, in a 45 Ti rotor for 40 min, yielding ~80 ml of a clear, yellow supernatant. This supernatant was applied to a 5 ml HisTrap column equilibrated with 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5 mM MgCl$_2$, 20 mM imidazole and 5% glycerol running at 1 ml min^{-1}. The column was washed with increasing concentrations of imidazole (20, 40 and 100 mM) in lysis buffer and protein was eluted with 250 mM imidazole in lysis buffer. Yellow fractions (typically three fractions or 6 ml) were pooled and concentrated to a volume of ~1.5 ml in a 30,000-MWCO spin concentrator (Amicon, Millipore) by centrifugation at 4 °C, 3,000g_{max}, in an Eppendorf centrifuge equipped with an A-4-44 rotor. The concentrated protein solution was applied to a 5-ml HiTrap desalting column (GE Healthcare) equilibrated with 25 mM Tris-HCl (pH 7.5), 50 mM NaCl and 5 mM MgCl$_2$. The protein was eluted with ~2 ml of this buffer.

The protein concentration of the solution from the desalting column was determined as described below and FAD was then added in at least fivefold molar excess from a 25 mM stock solution in 20 mM Tris-HCl (pH 7.5). To cleave off the His-tagged MBP, factor Xa (1 mg ml^{-1}, New England Biolabs) was added in a ratio of 0.5 mg per 100 mg of fusion protein. The protein solution was incubated at 4 °C in the dark for ~12 h and then diluted to ~1 mg ml^{-1} with 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5 mM MgCl$_2$, 20 mM imidazole and 5% glycerol and applied to the 5-ml HisTrap column equilibrated with the same buffer, to remove His-tagged MBP and uncleaved fusion protein. Washing with the same buffer was continued until the yellow peak had completely passed the column. The yellow fractions of the flow-through were pooled and concentrated in a 30,000-MWCO spin concentrator (Amicon, Millipore) as described above. The concentrated protein solution was injected onto a Superdex 200-pg 26/60 gel filtration column (volume, 330 ml) equilibrated with 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5 mM MgCl$_2$, 2 mM EDTA, 2 mM DTE and 5% glycerol. Gel filtration was performed at a flow rate of 1 ml min^{-1} and a maximum pressure of 0.5 MPa, and the purity of the yellow fractions was determined by SDS–PAGE. Pure yellow fractions were then pooled and concentrated in a

30,000-MWCO spin concentrator (Amicon, Millipore) by centrifugation at 4 °C as described above. The concentrated protein solution was applied to a 5 ml HiTrap desalting column (GE Healthcare) equilibrated with storage buffer (25 mM Tris-HCl (pH 8.0), 40 mM NaCl, 5 mM MgCl$_2$, 2 mM EDTA, 2 mM DTE, 5% glycerol). The column was washed with storage buffer and fractions containing the protein were collected. The BlrP1 protein concentration was determined optically using a Bradford assay and the FAD concentration was determined from its absorption at 450 nm.

The protein is >95% pure and contains only FAD, as judged by SDS–PAGE and HPLC analysis, respectively. Upon blue-light illumination, the absorption spectrum shifts ~10 nm to the red; the dark state recovers in ~1 min.

The protein solution (~350–400 μM or 15–20 mg ml^{-1}) was divided into small aliquots (20–50 μl) in brown tubes, flash frozen in liquid nitrogen and stored at −80 °C.

Site-directed mutagenesis of BlrP1. Mutations were introduced by Quick Change PCR mutagenesis using the following primers: BlrP_S309C–S312C_fwd, 5′-GGCGCAGGTTAT<u>TGC</u>GGCCTG<u>TGC</u>TTACTGACC CGT-3′; BlrP_S309C–S312C_rev, 5′-ACGG<u>GTC</u>AGTAAG<u>CACA</u>GGCC<u>GCA</u>ATAACCT GCGCC-3′. The L128C–G353C double mutant was generated by two subsequent Quick Change replacements using the following primers: L128C_fwd, 5′-GACCGGATGTTCCGGT<u>GCC</u>TTAGCGGCCTTCATT-3′; L128C_rev, 5′-AA TGAAGGCGCTAAGG<u>CACCGG</u>AACATCCGGTC-3′; G353C_fwd, 5′-TGCTG CGAGGATCTTT<u>GC</u>ATTACGGTGGTGGCT-3′; G353C_rev, 5′-AGCCACCA CCGTAAT<u>GC</u>AAAGATCCTCGCAGCA-3′. The substituted bases are underlined and shown in bold. All constructs were confirmed by DNA sequencing. The mutant proteins were expressed and purified as described for wild-type BlrP1.

The mutants were designed with the goal of introducing two adjacent cysteines, one on each BlrP1 monomer, that might form disulphide bridges and thereby sterically 'freeze' structural elements involved in signal transduction. However, only a small fraction of the protein was found to be crosslinked, whether the protein was purified in the presence or absence of a reducing agent (DTE). Addition of oxidizing agents increased the desired crosslinking only slightly (for example using glutathione), or resulted in multiple unwanted intermolecular cysteine crosslinks (for example using copper phenanthroline). Despite the fact that the desired disulphide bridges were not formed, the introduced bulky cysteine residues perturbed the integrity of targeted regions (compound helix, inter- and intramolecularly and the interfaces between the C-terminal BLUF helix and the EAL α6$_{EAL}$–β7$_{EAL}$ connector). The predicted effects of these perturbations on the enzymatic activity are consistent with the proposed model of intermolecular signal transduction.

Oligomeric state determination of BlrP1. BlrP1 (at both 1 and 8 mg ml^{-1} in storage buffer) was analysed in a Beckman Coulter Proteomelab XL-1 analytical ultracentrifuge using an An-60 Ti rotor at 72,600g_{max} and 20 °C in a two-sector centre piece with sapphire windows. Absorption scans were taken at 370 nm. Sedimentation velocity data were analysed using SEDFIT[29] (see also http://www.analyticalultracentrifugation.com), which showed BlrP1 to be a dimer in solution.

For analysis by dynamic light scattering with a Viscotek 802DLS instrument, the protein was diluted to 2 mg ml^{-1} in 50 mM Tris-HCl (pH 7.5). Care was taken to limit the amount of light reaching the experimental set-up.

Purification of YcgF. For purification of YcgF, we used DH5α expressing a GST–YcgF fusion from plasmid pGORF403 (ref. 12). Pelleted cells (42 g) obtained from a 30-l culture induced overnight at 18 °C were resuspended in 100 ml GSH buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM Na-EDTA, 2 mM DTT, 5% glycerol) and a tablet of Complete EDTA-free protease inhibitor cocktail (Roche) was added. The cells were lysed by three passes through a microfluidizer at a pressure of 0.8 MPa. Thereafter, the cell lysate was kept in the dark, and all manipulations were conducted in the dark or under red light at 4 °C. The lysate was clarified via ultracentrifugation at 142,000g_{max} in a 45 Ti rotor for 35 min at 4 °C. The supernatant was applied to a GSH column equilibrated with GSH buffer and washed overnight with GSH buffer. The GST tag was cleaved with thrombin (160 U) for 6 h at 4 °C, and cleaved GST removed by reapplying the eluate to the column using a peristaltic pump.

The eluted protein was incubated with 20-fold molar excess FAD and concentrated to a volume of ~7 ml in a 30,000-MWCO spin concentrator (Amicon, Millipore) by centrifugation at 4 °C, 3,000g_{max}, in an Eppendorf centrifuge equipped with an A-4-44 rotor. The concentrated protein solution was injected onto a Superdex 75 26/60 gel filtration column (volume, 300 ml) equilibrated with 25 mM Tris-HCl (pH 8.0), 300 mM NaCl, 2 mM Na-EDTA, 2 mM DTE and 5% glycerol. Gel filtration was performed at a flow rate of 1 ml min^{-1} and a maximum pressure of 0.5 MPa, and the purity of the yellow fractions was determined by SDS–PAGE. Pure yellow fractions were then pooled and dialyzed against 25 mM Tris-HCl, 40 mM NaCl, 2 mM DTT, 2 mM EDTA and 5% glycerol and

concentrated in a 30,000-MWCO spin concentrator (Amicon, Millipore) by centrifugation at 4 °C, as described above, to 1.3 ml, with a concentration of 20 mg ml^{-1}. The protein solution was divided into small aliquots (20–50 μl) in brown tubes, flash frozen in liquid nitrogen and stored at −80 °C.

Phosphodiesterase activity assays. Activity measurements were performed essentially as described previously[17]. However, because it became apparent that BlrP1 tends to stick to Eppendorf tube walls (to a lesser degree to low-affinity tubes than to regular ones), tubes were rinsed with 1 mg ml^{-1} BSA solutions just before the reaction solutions were mixed. Briefly, 140 μl of freshly mixed reaction mixture consisting of 1.1 μM protein (in 50 mM Tris-HCl (pH 7.5) and 50 mM NaCl) (YcgF or BlrP1) and 100 μM c-di-GMP were placed atop a 455-nm collimated LED light source at 20 °C (LEDC1, 22.3 mW cm^{-2} according to the manufacturer (THORLABS); 16 mW cm^{-2} according to measurements with a Gentec-EO Solo Power meter (Laser Components) attached to a PH100-Si photo detector; the reference wavelength was set to 450 nm, a region where the response curve of the detector is fairly flat).

The hydrolysis reaction was started by adding 10 mM MgCl$_2$ and allowed to proceed at 20 °C. Aliquots of 20 μl were taken at appropriate times and quenched by adding 10 μl 100 mM CaCl$_2$. Immediately afterwards, this mixture was boiled for 3 min and centrifuged. Then 20 μl of the supernatant was subjected to analysis by reversed-phase HPLC (buffer A: 100 mM KP$_i$ (pH 5.9), 4 mM TBAB; buffer B: 75% buffer A and 25% methanol). The pH dependence and divalent metal cation dependence assays were performed analogously with the exception that 50 mM Tris-HCl variously with pH 8.0, pH 8.5 and pH 9.3 was used instead of 50 mM Tris-HCl with pH 7.4, and that 10 mM CaCl$_2$ or MnCl$_2$ was used instead of 10 mM MgCl$_2$. In the case of Mn^{2+} as catalytic ion, the reaction was quenched with 3 μl 10% TCA because an initial burst of product formation hinted at insufficient displacement of Mn^{2+} by Ca^{2+}, and therefore incomplete quench of the reaction. Catalytic rates were obtained by linearly fitting the temporal decrease and increase in c-di-GMP and pGpG concentrations, respectively.

Crystallization and preparation of different metal complexes. Crystals of BlrP1 were grown by the hanging-drop vapour-diffusion method at 4 °C in the dark by mixing 1 μl of protein (19 mg ml^{-1}, in a buffer consisting of 5 mM c-di-GMP, 5 mM CaCl$_2$, 25 mM Tris-HCl (pH 8.0), 40 mM NaCl, 5 mM MgCl$_2$, 2 mM EDTA, 2 mM DTE and 5% glycerol) and reservoir (0.1 M Na-acetate buffer (pH 4.5), 0.2 M Ca^{2+}-acetate, 17–20% polyethylene glycol 3350) solutions. All experiments were set up and evaluated at 4 °C under red light, and crystallization plates were stored wrapped in aluminium foil. Crystallization plates were inspected with a 2-mm-thick RG630 filter (ITOS) shielding the microscope bulb. Crystals were rinsed in cryoprotection solution (0.1 M Na-acetate buffer (pH 4.5), 0.2 M Ca^{2+}-acetate, 23% polyethylene glycol 3350, 15% ethylene glycol) before flash-cooling in liquid nitrogen. To remove/substitute the bound calcium ions, crystals were soaked for 12–24 h at 4 °C in a solution containing 20 mM Na-EDTA (pH 7.0), 20 mM Na-EGTA, 0.1 M MES buffer (pH 6.0), and 17–20% polyethylene glycol 3350. The crystal used for the data collection of the metal-free complex was rinsed in EDTA/EGTA soaking solution augmented with 15% ethylene glycol and cryocooled in liquid nitrogen. To generate the Mn^{2+}-containing complexes, EDTA/EGTA-treated crystals were soaked for 0.5–5 min in cryoprotection solutions containing 50 mM Na-acetate and either 100 mM Na-MES (pH 6.0), 100 mM Tris-HCl (pH 8.0) or Tris-HCl (pH 9.0) and 50 mM MnCl$_2$.

Structure determination. All diffraction data were collected at beamline X10SA of the Swiss Light Source in Villigen, Switzerland, with the crystal kept at 100K and data processed with XDS[30]. A single-wavelength anomalous diffraction data set and the 2.4 Å resolution calcium complex data set collected at pH 4.5 were

used to determine the structure with AUTOSHARP[31]. The anomalous phasing power was 0.733, and the figure of merit for acentric reflections after SHARP phasing was 0.269. Solvent flattening by AUTOSHARP yielded an excellent electron density map. The largest part of the structure was then built automatically using a script obtained from the RESOLVE website (http://www.solve.lanl.gov/Resolve/resolve.html), which cycles density modification and automatic building using RESOLVE[32–34] with refinement by REFMAC5[35]. In this way, >90% of the main chain was built, and >50% of the sequence was docked into the density. The remainder of the model was built using XTALVIEW/XFIT[36] and COOT[37]. Simulated annealing was performed with CNS[38,39], and the final model was obtained by repeated cycles of rebuilding in COOT and refinement with REFMAC5. Refinement statistics are given in Supplementary Table 1. Despite the relatively high resolution of 2.4 Å, NCS restraints had to be applied to the core of the molecule to allow refinement to a low R_{free} factor. Into each monomer, a c-di-GMP molecule, a calcium ion and a flavin molecule were built. The latter was modelled as an FMN molecule because there was no electron density for the solvent-exposed AMP part of FAD, similarly to the situation in other BLUF structures[2,21].

For the other complex structures, a BlrP1 model from which all water molecules, the metal ions, and the c-di-GMP- and FMN molecules were removed was refined against the respective complex crystal data using REFMAC5[35], and any ligands clearly shown to be present by the sigma-a weighted double-difference and difference electron density maps were built in. No restraints were placed on the metal–protein interaction geometry. The use of NCS restraints, TLS parameters and riding hydrogen atoms was guided by the R and R_{free} values. Residues with poor backbone geometry as indicated by the Ramachandran plot are in flexible regions as shown by poor electron density. In the final models, >99% of the residues are in the core and allowed regions of the Ramachandran plot. Figures were generated with PYMOL[40].

29. Schuck, P. & Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54**, 328–341 (2000).

30. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–806 (1993).

31. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).

32. Terwilliger, T. C. Automated structure solution, density modification and model building. *Acta Crystallogr. D* **58**, 1937–1940 (2002).

33. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).

34. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).

35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

36. McRee, D. E. XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).

37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

38. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

39. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).

40. DeLano, W. L. *PyMOL Molecular Viewer* ⟨http://www.pymol.org⟩ (2002).

NEWS

Women at the top

Women faculty members in science, maths and engineering are climbing the academic ladder at top US research universities, according to a report released on 2 June by the US National Research Council. The study finds that women do well once they are in the pipeline for academic opportunities, but that they are still under-represented in the top ranks.

"There's great equity in terms of job interviews, hiring and promotion," says Sally Shaywitz, co-chair of the committee that wrote the report and co-director of the Yale Center for Dyslexia and Creativity in New Haven, Connecticut.

Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty looks at how women fare at key points in their academic careers compared with men. The congressionally mandated study surveyed almost 500 departments and more than 1,800 faculty members at 89 research-intensive universities in 2004 and 2005. It focused on six disciplines: biology, chemistry, civil engineering, electrical engineering, maths and physics.

A higher percentage of women applying for their first job at major research universities get interviews and receive offers for tenure-track positions than their representation in the candidate pool would predict. But many women PhDs are not applying for these positions. Between 1999 and 2003, women earned 45% of biology PhDs, but comprised 26%

Sally Shaywitz.

of the applicants for faculty positions at the universities surveyed.

Men and women faculty members spent similar amounts of time on teaching, research and service, and there was little difference in the number of refereed publications, grant funding or award nominations between the two groups. The salaries of male full professors averaged about 8% higher than female full professors', but that disparity could be the result of differences in seniority, says the report.

Representatives of women researchers are more critical. "If you're only looking at elite research universities with the very top women being recruited, you will see relatively few gender-specific inequities in appointments and promotions. To conclude that there are no problems is unwarranted," says Phoebe Leboy, president of the Association for Women in Science and professor emerita of biochemistry at the University of Pennsylvania in Philadelphia.

That point was made by physicist Claude Canizares, committee co-chair and vice-president for research at the Massachusetts Institute of Technology in Cambridge. He called for longitudinal studies on why women do not apply for faculty positions at research-intensive universities. "We studied a subset of women — the ones who do make it," he says. "We don't know anything about the women who left." ■

Hannah Hoag is a science journalist based in Montreal, Canada.

POSTDOC JOURNAL

Befriending rejection

A particularly memorable scene from the film *Apocalypse Now* — set in the Vietnam War — features Colonel Kurtz, played by Marlon Brando, lazily articulating the line: "You must make a friend of horror." When I become wealthy and remake this film, I will set it in academia. In my version, Professor Kurtz says "You must make a friend of rejection."

Rejections take a variety of forms, from impersonal platitudes in a form letter to seemingly personal comments in a review. Since graduating from college, I

have received many such rejections. At first, I took them personally. I wrote (but ultimately did not send) a vitriolic letter in response to the review that sunk my first paper. The year I had five different papers rejected was a low point, a time when I questioned whether I could continue in academia. But now, after I've eked out a few acceptances, I see that rejection is simply part of academia.

I recently started receiving responses from my faculty applications, and the hours of work I spent on those

applications are being dismissed by a few sentences in a letter or a few lines in an e-mail.

But, just as negative reviews have strengthened my scientific writing, so will these rejections strengthen my job application. I will reword, rewrite and, above all, keep publishing. Then, I propose with confidence, I will find my job. Even without the advice of Professor Kurtz, I have made a friend of rejection. ■

Sam Walcott is a postdoc in theoretical biophysics at Johns Hopkins University in Baltimore, Maryland.

IN BRIEF

Biosafety training

The University of Texas Medical Branch in Galveston has received US$5 million in federal funds to expand its biosafety-training programme and facility. The university's course has trained 1,200 infectious-disease researchers and others since it started in May 2005.

The new training centre, one of only a few in the United States, is slated for completion within two years. It will include a more comprehensive biosafety level-4 component as well as more extensive facilities, and will offer training for biosafety support staff and engineers as well as researchers. Trainees from outside the federal government are eligible to apply.

Science for physicians

Medical education in the United States should place more emphasis on basic scientific competence, says a report published this month.

Scientific Foundations for Future Physicians, by the Howard Hughes Medical Institute and a committee of the Association of American Medical Colleges (AAMC), warns that pre-med and medical-school curricula have not kept pace with advances in scientific knowledge. Areas for improvement include a better understanding of the natural sciences in fields such as the biology of microorganisms and the principles of pharmacology. The AAMC is studying the report as part of a review of its medical college admission test.

Venezuela science boost

The Venezuelan government announced on 3 June a grant of 336 million bolivars (US$156 million) to the country's research universities, with the largest amount, 66 million bolivars, going to the Central University of Venezuela in Caracas. Science and technology investment has risen in recent years to 2.69% of gross domestic product, said science minister Jesse Chacón. The number of research scientists working in Venezuelan universities more than tripled between 1999 and 2008, from 1,689 to 6,038, he said. However, government-funded institutions recently had their budgets cut by 6%, although salaries were not affected (see page 898).

TEXAS-SIZED CHALLENGE

Houston's Texas Medical Center is a biomedical behemoth. But it's not immune to the souring economy, as **Gene Russo** finds out.

Postdoc Mala Pande is just one academic victim of the credit crunch. A postdoc at the University of Texas M. D. Anderson Cancer Center in Houston, studying how gene–environment interactions lead to cancer, her upcoming search for a permanent position just got a lot harder. In March, M. D. Anderson imposed a hiring freeze, as have other academic institutions throughout the United States (see *Nature* **458,** 372–373; 2009). Although Pande would prefer to stay in Houston, she has had to expand her job search to other regions — and even outside the United States — to any institution that specializes in her field. "I get a sense there are jobs, but then there are all these hiring freezes," says Pande, adding that she feels more pressure to publish and to write grants for funding from outside her institute to retain a competitive edge. Her story exemplifies the current financial difficulties that are affecting even world-renowned institutions in the United States.

M. D. Anderson is part of the Texas Medical Center (TMC), a vast assortment of hospitals and labs located on an area of 400 hectares south of downtown Houston, and one of the largest medical centres in

"I get a sense there are jobs, but then there are all these hiring freezes."
— Mala Pande

the world. The numbers are impressive: 47 member institutions, including 19 academic ones, boasting between them 15,400 physicians, researchers and other professionals, and 33,000 students from high school to graduate level (see page 896).

But even the TMC has been hit by the economic downturn. In a mid-April statement, M. D. Anderson president John Mendelsohn warned that the institute's growth in expenses had outpaced growth in revenue for six consecutive months. Revenue had dropped in part because of an increase in the number of uninsured patients. If M. D. Anderson takes no action, revenues will grow at a slower pace than costs. "We want to make sure we're in good financial strength before hiring," says Raymond DuBois, M. D. Anderson's provost and executive vice-president. "And this year, we had already hired more than we'd planned." M. D. Anderson is, however, hiring a limited number of 'mission critical' employees. There are fewer than 70 positions open, down from 1,000 in March.

Other institutions in the region are in even worse financial straits. Unlike M. D. Anderson, next-door neighbour Baylor College of Medicine has no hospital to generate income; a 'Baylor Hospital and

Clinic' is under construction, but a lack of funds means that, for now, Baylor has no plans to move in individuals or equipment, only to finish the exterior. Baylor's endowment has dropped by 30% because of the downturn, the college reports, and the annual operating revenues derived from its endowment have fallen by about $16 million. There are ample research areas for investment, notes Susan Hamilton, senior vice-president and dean of research at Baylor. "But we have to get past financial challenges before we can expand too much." She adds that Baylor is still hiring in "strategic areas" that include McNair Scholars, a programme that seeks to support rising stars in biomedicine. And the college is in discussions with Rice University in Houston to complete a merger; Baylor would get the backing of a major university and Rice would get a renowned medical school. But so far, the two institutions have only a memorandum of understanding.

Cautious growth

Nevertheless, despite hard economic times, the TMC is still expanding. Seven hundred thousand square metres of new build is under construction, at a cost of $3.3 billion, and one-third of this space is designated for research. Baylor recently completed the Margaret M. Alkek Building for Biomedical Research, which cost $100 million. Four floors, each of 1,850 square metres, house interdisciplinary research space focusing on cardiovascular sciences, diabetes, cancer, pharmacogenomics, imaging and proteomics.

Helen Shepherd, Baylor's director of research infrastructure programmes, says that the new building is part of an effort to promote interdisciplinarity. Meanwhile, Baylor, in addition to its hospital project, is working with Texas Children's Hospital to complete a neurosciences institute, although a funding shortfall means that some floors will be left vacant and partially built for now. At M. D. Anderson's south campus, some floors of a new advanced biomedical imaging centre — a collaboration with the University of Texas Health Science Center — may also be left incomplete.

Still, efforts to forge interdisciplinary collaborations continue. The newest addition to the landscape is Rice University's Bioscience Research Collaborative (BRC), due to welcome its first Rice faculty and external tenant on 1 July. The 10-storey building houses open-plan, state-of-the-art interdisciplinary laboratory space as well as two auditoria, and has plenty of room for core facilities and animal housing. Open-plan lounges for graduate students and postdocs — designed to encourage collaboration — overlook the manicured grounds of the university.

Tough times

But finding tenants for the BRC has not been easy. Rice built the facility in the hope of promoting collaboration and cooperation with its immediate neighbours, but the only confirmed external tenant so far is the Gulf Coast Consortium, a coalition of six Houston-area research institutions aiming to build interdisciplinary collaborative research teams and training programmes, which is renting office space there. Jim Coleman, Rice's vice-provost for research, says potential tenants such as Baylor initially expressed interest, then decided against. "Economically it's a very odd time for all of us," says Shepherd.

There's also the issue of whether researchers will find the BRC set-up agreeable. Unlike most interdisciplinary open-lab forays, the BRC designers wanted labs from different institutions to set up next to each other. This could prove uncomfortable if neighbours are competing for the same grants.

Newly recruited Rice University professor and BRC occupant John McDevitt isn't worried on that score. "For the most part, the people that decide to come together in this kind of joint interaction are very much developing a co-dependence on one another," he says, adding that he thinks

Rice University's new Bioscience Research Collaborative provides space for interdisciplinary research, but Jim Coleman (inset) says tenants are hard to find.

"We have to get past financial challenges before we can expand too much."
— Susan Hamilton

Rice will select tenants carefully. For McDevitt, the BRC is a dream situation. He is joining the facility after spending 20 years at the University of Texas at Austin, where he has been refining his special brand of miniaturized 'lab on a chip'. One of his chips, for example, can be used to rapidly measure the degree of immunodeficiency in HIV-infected individuals, and so gauge their progression towards AIDS. McDevitt hopes the chips will be of use in remote regions of Africa where more conventional and bulkier assay methods cannot go.

At the BRC, McDevitt will be able to draw not only on Rice's world-class nanotechnology expertise but, most crucially, he will have better access to samples from patients at the nearby hospitals. He has already established a collaboration with Christie Ballantyne, a professor of medicine at Baylor, as part of an effort to develop a chip that serves as a rapid screen for heart attacks. McDevitt is bringing 25 students, postdocs and other scientists with him from Austin. "I feel much better being at the BRC than being at a medical school," he says. "We really need the clinical piece, but we also need strength in science, engineering and nanotechnology."

Despite the vast expanse of biomedical expertise in the TMC, Houston's biotech sector is not a huge generator of jobs, especially at entry level. There are roughly 160 biotech, device and diagnostics companies in and around the TMC, three times the number six years ago, but more than half of

the firms have fewer than 10 employees. And many companies are struggling financially or are more likely to seek candidates with mid-level or higher skills and experience, according to Jacqueline Northcut, chief executive of Biohouston, a non-profit organization founded by TMC institutions to raise the region's prowess in biotech. One potential shot in the arm could come with a significant wedge of state funds for industry and academic cancer research as part of the Cancer Prevention Institute of Texas (CPRIT) initiative. First approved in 2007, the institute received $450 million in funding through Texas state legislative approval early this month. But if CPRIT, inspired by the California stem-cell institute, gains its goal — $300 million in state cancer research funds per year for 10 years — that should have a big impact on cancer researchers in academia and in pharmaceutical and biotech companies.

For now, searching for a biotech job can be discouraging, says M. D. Anderson postdoc Ivone Bruno. Head of the postdoc association there, Bruno had been expecting to stay at M. D. Anderson in a full-time 'instructor' position. But the verbal offer fell through after the hiring freeze. Like Pande, Bruno does not want to relocate — she likes the area, and her high-school-aged daughter wants to avoid switching schools just as she is about to apply for college. Bruno would be interested in a biotech position, but work is scarce. She is trying to stay positive; the lack of work gives her an opportunity to expand her skills — she has already moved from molecular biology to neurobiology to stem-cell biology — and explore other research possibilities.

"This lets me go beyond my comfort zone," Bruno says. She and other Houston-area jobseekers will probably have to continue to seek the silver lining in a difficult situation. ∎

Gene Russo is editor of *Naturejobs*.

Totipotent

Caught in a trap.

Catherine Krahe

One cell is not enough to speak or think. Perception is intermittent and mostly chemical. But one cell can remember.

The plantlet sits on a culture plate with a dozen dying samples. It has been one cell, and each day brings a new memory to the surface. It remembers the knives and the flames. It remembers the pin guns and such pain. It remembers Bea, the hands the scent the voice, and it remembers Bea, the Monkey, who tore him to pieces and left him behind, and he remembers Bea, his partner, who scraped green from under her fingernails and walked out of the lab.

He remembers Bea.

He has ends now, root and shoot. He has differentiated cells. He does not look like the Flytrap he is, but he can perceive.

There is light, and there is Bea. She smells cold, afraid, fierce. Someone moves him to a new dish.

One cell cannot act; a mass of photosynthetic tissue can hope. Mimosa, Mosy, remembers.

He remembers talking. He remembers Bea, Bea panting, Bea bleeding, Bea becoming still while her scent changed from tenacity to bewilderment and sorrow. He remembers … he remembers fighting while Bea ran for the lifeboat. He remembers pin guns and how his partner would die forever.

He has leaves now. He has appendages. He looks like a Flytrap.

Bea is alive. Mosy can hear her again.

"Hey." Mosy can't answer; this part of him doesn't have a mouth yet. "They say you're gonna pull through." A gripping tendril curls towards the Monkey. *His* Monkey. "I talked with the rest of you. You wanted to be bigger before you came back. You'll tell me when you're ready, right? It's not the same talking to the rest of you. They weren't there. You, I mean."

Monkeys do not live like Flytraps. A Flytrap cut in two feels two wounds, roots starving and sending up leaves, leaves thirsty and wilted, struggling to make roots. They remain separate, yet think and act as one. Mimosa has three greenhouses on three planets; the dozen cells that never multiplied are no loss. He remembers all that he does, and once he is near the rest of him, so will he. Like all Flytraps, he sends cuttings, plantlets filled with knowledge, to each greenhouse, spreading the word, the thoughts, the memories.

Bea is afraid of the greenhouses now. Mosy can hear her voice shake when she talks about the part of him that lives there.

Mosy has a pot, heavy white and black ceramic. Bea brought it to tell him they're not moving. The Monkey carries him to his private greenhouse. Halfway there, he discovers and remembers the months of waiting and months in gel, months in space and glass and multiple homecomings. He screams and the plantlet in Bea's hands hears it vibrate through the hall. He remembers sending Bea away after two uncomfortable visits and understands both why he did so and why it was so cruel. The plantlet grown from a few battered cells has become part of Mosy. Bea puts the pot on a table and sits in her chair by the frost-etched glass.

Mosy wants to touch her from all sides. Bea flinches away, but reaches for the crooked plant in the white and black pot.

"I don't want you going out there again," Mosy says. His partner jumps.

"Why not? You told me to leave before."

He remembers this. He thought it would be good for the unhappy Monkey. Now he remembers the fighting and flushes purple at the danger. The memory — hiding, both wounded, Bea's confession — Mosy thinks of thorns and cells divide within him.

"If you die, you die forever," he says. The confession during the fight. He will hold Bea here if he has to.

"That's the usual way of it."

Mosy shudders; Bea doesn't react as the greenhouse trembles. "Please don't. I couldn't — it will hurt."

"Mosy —" Bea stands and pets a tendril. She tastes warm and upset.

"If you die you die forever. And I'll have to know. I'll have to tell the rest of me how it happened. I'll find out again and again."

"I have to stay here so you don't cry?" Bea moves further into Mosy's greenhouse. "Death happens no matter what. You can have me on a ship with you or alone —"

"With me down here!"

"— alone, because I'm not staying! We're partners, and that means we go up together. Or we go alone. And Monkeys die."

She leaves. Mosy shakes for hours. He doesn't know how to tell the rest of him. He doesn't know how to keep Bea safe. Monkeys die forever. Bea will die forever.

If he does not touch the windows, he can pretend it isn't cold outside. The greenhouse stays warm long after dark. It is no comfort. ∎

Catherine Krahe is an environmental engineering student at the University of Iowa who plans to save the world by telling stories and planting trees.

FUTURES